

Le modèle de l'alignement curriculaire élargi pour étudier les pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire: enjeux conceptuels et pragmatiques

PASQUINI, Raphaël Giuseppe

Abstract

Cette recherche doctorale documente d'une part les mouvements dans les pratiques évaluatives sommatives d'enseignants en exercice lorsqu'elles sont confrontées, au sein d'un dispositif de formation, au modèle théorique de l'alignement curriculaire élargi, et d'autre part, l'évolution conceptuelle du modèle lorsqu'il est mis à l'épreuve des pratiques réelles et des contextes des participants. Pour documenter ces mouvements et voir dans quelle mesure les pratiques d'évaluation pouvaient gagner en cohérence et ainsi servir les apprentissages, nous avons catégorisé puis triangulé quatre sources de données (des entretiens compréhensifs, des écrits exemplaires, des épreuves sommatives réalisées par les participants et les échanges en séminaires de formation). Il ressort notamment de nos résultats que si le modèle est un outil théorique compréhensif puissant pour comprendre les pratiques d'évaluation voire les amener vers plus de cohérence, les compétences évaluatives des enseignants ne se développent pas de manière linéaire, mais bien davantage en fonction des contextes et des connaissances [...]

Reference

PASQUINI, Raphaël Giuseppe. *Le modèle de l'alignement curriculaire élargi pour étudier les pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire: enjeux conceptuels et pragmatiques*. Thèse de doctorat : Univ. Genève, 2018, no. FPSE 704

DOI : 10.13097/archive-ouverte/unige:106442

URN : urn:nbn:ch:unige-1064421

Available at:

<http://archive-ouverte.unige.ch/unige:106442>

Disclaimer: layout of this document may differ from the published version.



UNIVERSITÉ
DE GENÈVE

TITRE DE LA THESE

Le modèle théorique de l'alignement curriculaire élargi pour étudier des pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire : enjeux conceptuels et pragmatiques

THESE

Présentée à la
Faculté de psychologie et des sciences de l'éducation
de l'Université de Genève
pour obtenir le grade de Docteur en **Sciences de l'Éducation**

par

Raphaël PASQUINI

De Lugano (TI)

Thèse No 704

GENEVE

11 Juin 2018

Numéro d'étudiant : 84-807-114

Membres du jury

Lucie Mottier Lopez (directrice), Faculté de psychologie et des sciences de l'éducation, Université de Genève

Stéphane Clivaz, UER de didactique des mathématiques et des sciences de la nature, Haute Ecole Pédagogique du canton de Vaud, Lausanne

Valérie Lussi Borer, Faculté de psychologie et des sciences de l'éducation, Université de Genève

Christophe Ronveaux, Faculté de psychologie et des sciences de l'éducation, Université de Genève

Walther Tessaro, Faculté de psychologie et des sciences de l'éducation, Université de Genève

Catherine Van Nieuwenhoven, Faculté de psychologie et des sciences de l'éducation, Université catholique de Louvain

Reconnaissance

Nicolas Bouvier, dont les écrits ont marqué ma vie de jeune adulte, doutait des gens affirmant « ne devoir rien à personne ». J'ai envie ici de saluer toutes celles et ceux qui m'ont accompagné lors de cette thèse et envers qui j'ai quelques dettes...

Je peine à trouver les mots qui rendraient justice à la qualité de la direction de Lucie Mottier Lopez. Depuis notre premier entretien, elle m'a offert sans compter et avec bienveillance son temps, ses compétences, ses réseaux. La rigueur de sa pensée, son niveau d'exigences, son sens du débat et son acuité sur les dimensions scientifiques de l'évaluation et de la recherche ont contribué à faire de cette thèse une aventure intellectuelle fondatrice.

Merci à Stéphane Clivaz, didacticien des mathématiques, d'avoir participé au jury de cette thèse. Je le comprends comme la suite d'une collaboration qui a pris forme il y a déjà dix ans dans la Broye, puis en Asie.

De nombreuses cogitations méthodologiques et théoriques ont parsemé cette période de quatre ans. Merci à Valérie Lussi Borer de les avoir discutées avec ouverture et intérêt.

Le questionnement de Christophe Ronveaux m'a permis de mieux entrer dans les dimensions didactiques de l'évaluation en français. Un grand merci à lui pour ses apports.

Spécialiste reconnu du champ de l'évaluation, je suis heureux que Walther Tessaro ait accepté de faire partie de la commission et du jury : son expertise a su participer à la complétude de cette thèse, un immense merci à lui.

La dimension collaborative de cette recherche a pris de l'ampleur tout au long de ce projet. Merci à Catherine Van Nieuwenhoven pour ses connaissances partagées avec autant de passion.

Mes remerciements chaleureux vont aussi aux enseignants qui se cachent derrière les prénoms d'emprunt. Leur engagement et la relation de confiance qui nous a unis ont permis à cette recherche d'être également pour moi une magnifique expérience de partage avec des pédagogues.

Merci encore aux membres du Comité de Direction de la HEP Vaud, Guillaume Vanhulst, Cyril Petitpierre et Luc Macherel, de m'avoir accordé leur confiance. Les conditions offertes sont un

privilège et malgré un emploi du temps toujours trop chargé, elles m'ont permis d'avancer raisonnablement.

J'adresse ma profonde reconnaissance à mes responsables d'unité, Anne Clerc-Georgy et Daniel Martin, d'avoir valorisé ce travail. Qu'ils soient ici chaleureusement remerciés, ainsi que tous les collègues de mon équipe, pour leur intérêt envers ce que cette thèse m'a permis de transmettre dans les actions de formation traitant d'évaluation.

Merci à mon collègue de bureau et lui aussi doctorant, Lionel Dechamboux, pour toutes nos discussions épistémologiques, conceptuelles, méthodologiques, pour les remontées de moral, pour les (fou)rires, pour ce sentiment de ne pas être seul dans l'aventure.

Une pensée particulière va à Nicolas Perrin, incollable expert du logiciel « word ». Merci à lui de m'avoir évité de multiples crises de nerfs.

Je serai éternellement reconnaissant et admiratif envers ma femme, Sonya, de m'avoir accepté dans cette mouvance intérieure, alors que la vie nous appelait souvent avec force ailleurs. Son intelligence, sa créativité pour alléger parfois le quotidien, son empathie hors norme, ont été des respirations inestimables. Je prends également le parti de remercier ici mes enfants, Elise et Mathieu (nés juste avant et en plein processus d'écriture), d'avoir été des sources intarissables d'énergie et de confiance dans le monde.

Tant d'autres personnes méritent encore un merci. Elles ont été présentes, à leur manière, pour supporter, égayer, distraire, écouter, parler, soulager, nourrir, faire voyager : ma famille, mes amis, mes partenaires de tennis, mon masseur, mon acupunctrice, et toutes celles que j'ai sûrement oubliées.

Nicolas Bouvier avait raison : on finit par ne plus compter ses dettes humaines. Et c'est bien comme ça.

La surprise nous oblige à réviser nos systèmes d'explication qui ne l'avaient pas prévue, de penser à ce qu'elle signifie de nouveau, et éventuellement ce qu'elle annonce pour le futur.

Edgar Morin, *Au rythme du Monde*, 2015

La signature du regard est une affaire intime.

Raymond Depardon, *Errance*, 2004

Table des matières

Partie 1 - Introduction.....	19
1 Prolégomènes.....	19
2 Un parcours parsemé d'interrogations	19
2.1 Les premières questions de l'enseignant	19
2.2 Les questions du formateur.....	20
2.3 Des questions pour le chercheur.....	26
3 Les enjeux relevés par la thèse	28
4 Organisation du manuscrit	28
Partie 2 - Revue de littérature.....	31
1 Introduction.....	31
2 L'étude des pratiques d'évaluation sommative, ou l'expression de tensions.....	32
2.1 L'évaluation normative : une évaluation par défaut	33
2.2 Des pratiques évaluatives contrastées	33
2.3 De l'(in)cohérence et de l'(in)validité.....	35
2.4 Des contextes très influents	36
2.5 Les principales limites des recherches convoquées	36
2.6 Des travaux sur l'évaluation sommative issus des didactiques	37
3 La note : ou la nécessité de reformuler le débat qu'elle suscite	42
3.1 De l'urgence à (re)définir la finalité de l'évaluation sommative notée	43
3.2 Le problème : la construction de la note	44
3.3 Une note pour jouer trop de rôles	45
3.4 Une urgence supplémentaire : re-considérer la note en référence aux apprentissages	46
4 Un modèle pour appréhender la cohérence de l'évaluation sommative : l'alignement curriculaire	47
4.1 L'alignement curriculaire : un réseau conceptuel	47
4.2 Un modèle linéaire	48
4.3 Une taxonomie pour lire la cohérence d'alignement curriculaire	49
4.4 Une cohérence « <i>top down</i> » et « <i>bottom-up</i> »	51
5 Une acception de l'alignement curriculaire nécessairement élargie	52
5.1 Les conditions pour un curriculum à des fins d'alignement curriculaire	53
5.2 Le rôle clé des tâches évaluatives	56
5.3 Une littérature abondante sur la pondération.....	57

6	Une notation différente au travers de l'évaluation critériée	64
6.1	Une notation holistique qui questionne l'échelle ordinale traditionnelle	65
6.2	Une notation à visée qualitative : quelques enjeux	67
6.3	Des notes pour une évaluation sommative au service des apprentissages.....	68
6.4	Des constats fondamentaux sur une notation référée aux apprentissages	68
7	Des recherches pour une note alignée curriculairement	69
7.1	Une note sans mesure ?	69
7.2	Quelle échelle adopter ?.....	70
7.3	Quel barème élaborer dans ce cas ?	70
8	Au centre : une prise d'informations référées à un modèle théorique	71
8.1	Le jugement professionnel de l'enseignant : un geste omniprésent.....	72
8.2	Le jugement professionnel au service d'une cohérence d'alignement curriculaire ..	73
9	Pourquoi ne parler que de la validité de l'évaluation ?.....	74
10	Les paradigmes d'évaluation : des systèmes de production de jugements spécifiques	76
11	Quelques points de convergence pour une épistémologie de l'évaluation	78
12	Une revue volontairement orientée	78
Partie 3 - Cadre conceptuel relatif à l'étude des pratiques évaluatives des enseignants.		81
1	Introduction.....	81
2	Evaluation sommative ou certificative ?.....	83
3	Un modèle théorique phare : l'alignement curriculaire	84
3.1	Une première conceptualisation de l'alignement curriculaire incomplète	85
3.2	Une conceptualisation plus approfondie	88
3.3	Quelques incomplétudes théoriques relatives à l'évaluation sommative	89
4	Une nécessité théorique d'élargir le modèle : l'alignement curriculaire élargi.....	90
4.1	Un modèle devenant avant tout systémique	91
4.2	Un modèle qui met en évidence la qualité de l'apprentissage dans l'évaluation sommative	93
5	Le curriculum, une référence parmi bien d'autres	94
6	Les objectifs d'apprentissage, ou la question de la référenciation	95
6.1	Vers une référenciation restreinte	96
6.2	Des objectifs d'apprentissage orientés théoriquement.....	96
7	Les tâches évaluatives sommatives : une question de <i>design</i>	98
7.1	Des tâches authentiques	99

7.2	Quelles caractéristiques théoriques des tâches évaluatives ?	100
8	La pondération ou la question de la valeur de l'apprentissage	101
8.1	Les critères d'évaluation : l'unique outil pour mettre en valeur la qualité de l'apprentissage	102
8.2	La question du poids donné aux critères	106
9	La nécessité de conceptualiser différemment la notation	107
9.1	Noter et évaluer : quelle articulation ?	108
9.2	La notation comme un processus	109
9.3	Une notation, de fait, holistique	110
9.4	Des échelles holistiques ?	113
9.5	Le rôle des disciplines dans les processus de notation holistiques	113
10	Vous avez dit jugement professionnel ?	114
11	Et concevoir un jugement curriculaire holistique ?	115
12	Une recherche permanente de validité	116
Partie 4 - Cadre conceptuel sur l'<i>assessment literacy</i>		118
1	Introduction	118
2	L' <i>assessment literacy</i> : une perspective théorique pour comprendre et instrumenter le développement des compétences en évaluation	119
2.1	Comment comprendre <i>literacy</i> ?	120
2.2	Une littératie en évaluation ?	120
2.3	De quelles connaissances parle-t-on dans une perspective théorique d' <i>assessment literacy</i> dynamique ?	127
2.4	Qu'en est-il des compétences à évaluer en <i>assessment literacy</i> du point de vue de la recherche ?	128
2.4.1	Un premier ensemble de réflexions intéressantes pour une	128
2.4.2	Un second apport significatif sur une	129
2.5	Les compétences en <i>assessment literacy</i> retenues dans le cadre de notre recherche	130
3	Pour ou contre une <i>assessment literacy</i> en évaluation sommative ?	132
4	En guise de synthèse	133
5	Questions de recherche	134
Partie 5 - Présentation du dispositif de recherche		136
1	Introduction	136
2	Contexte de la recherche	136

2.1	Une première stratégie de recrutement d'enseignants infructueuse.....	136
2.2	Une nécessaire réorientation du recrutement.....	137
2.3	Des changements importants dans le choix des disciplines.....	138
2.4	Une campagne de recrutement difficile	139
3	L'entrée en matière avec les participants.....	140
4	Les profils des enseignants volontaires.....	140
4.1	Les quatre enseignantes de français	140
4.2	Les quatre enseignants de mathématiques	142
5	Le déroulement de la recherche.....	143
5.1	Un dispositif de recherche-formation dual	143
5.2	Les différents temps du dispositif de recherche-formation	146
6	En guise de transition	150
Partie 6 - Ancrage méthodologique.....		152
1	Introduction.....	152
2	Les raisons de mener une recherche collaborative.....	154
2.1	La genèse de la recherche collaborative	154
2.2	Les caractéristiques de la recherche collaborative	156
2.3	Les modalités spécifiques de participation des praticiens dans une recherche collaborative.....	162
2.4	Un rôle non moins spécifique pour le chercheur.....	163
2.5	Qu'en est-il du chercheur-formateur ?	164
2.6	Une perspective de recherche fondamentalement compréhensive et interprétative	166
3	L'univers interprétatif	166
4	Quatre sources de données.....	167
4.1	Des écrits pour accéder à une forme de prise de recul.....	167
4.2	Une deuxième source de données : les entretiens	170
4.3	Notre troisième source de données : des présentations et discussions d'évaluations sommatives lors des actions de formation	172
4.4	La quatrième source de données : des évaluations sommatives référées aux contextes de classe	173
5	Une démarche d'analyse dans un rapport dialectique entre déductif et inductif....	174
5.1	L'analyse à l'aide des catégories conceptualisantes	175
5.2	Catégorie et catégorisation	178

5.3	Un travail de déduction et d'induction, encore...	179
6	L'analyse des données : une nécessaire triangulation	179
6.1	Une stratégie de recherche particulière	180
6.2	Une triangulation de méthodes	181
6.3	Des perspectives de triangulation poursuivant des buts différents	181
6.4	Une première phase d'analyse à approfondir	183
6.5	Une seconde phase d'analyse interprétative et dynamique	183
7	Deux parties pour exposer les résultats	185
Partie 7 - Premier chapitre de résultats : six portraits d'enseignants		187
1	Introduction	187
2	Portrait 1 : Fernand, enseignant de mathématiques	188
2.1	Les dimensions des pratiques évaluatives de Fernand sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi	189
2.2	Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Fernand	212
2.3	En guise de synthèse provisoire	215
3	Portrait 2 : Paul, enseignant de mathématiques	216
3.1	Les dimensions des pratiques évaluatives de Paul sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi	217
3.2	Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Paul	240
3.3	En guise de synthèse provisoire	243
4	Portrait 3 : Philémon, enseignant de mathématiques	244
4.1	Les dimensions des pratiques évaluatives de Philémon sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi	244
4.2	Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Philémon	267
4.3	Le processus de développement des compétences évaluatives chez Philémon au regard des caractéristiques de l'assessment literacy	268
4.4	En guise de synthèse provisoire	269
5	Portraits 4 et 5 : Anouck et Alizée, enseignantes de français	271

5.1	Les dimensions des pratiques évaluatives d'Anouck et Alizée sur lesquelles elles portent chacune leur attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.....	271
5.2	Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez les deux enseignantes	307
5.3	Le processus de développement des compétences évaluatives chez chaque enseignante au regard des caractéristiques de l'assessment literacy	310
5.4	En guise de synthèse provisoire	311
6	Portrait 6 : Samia, enseignante de français	312
6.1	Les dimensions des pratiques évaluatives de Samia sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi	312
6.2	Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Samia	337
6.3	En guise de synthèse provisoire	340
7	Conclusion	341
Partie 8 - Second chapitre de résultats		345
1	Introduction.....	345
2	La description des pratiques de notation par les enseignants	346
2.1	Les phénomènes, questions et réflexions sur lesquels les enseignants portent leur attention dans leurs pratiques de notation avant la formation.....	347
2.2	Nature des obstacles rencontrés par les enseignants dans leurs pratiques de notation	352
2.3	Trois apports sur les pratiques de notation du point de vue de l'alignement curriculaire élargi.....	365
3	Conclusion	367
Partie 9 - Conclusion		370
1	Propos liminaires	370
2	Les principaux apports de notre recherche	371
2.1	Les dimensions des pratiques évaluatives sur lesquelles les enseignants portent leur attention lorsqu'elles sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.....	371
2.2	Les compétences évaluatives spécifiques d'un point de vue sommatif que l'alignement curriculaire élargi participe à développer.....	372

2.3	L'évolution des pratiques de notation chez les enseignants	374
2.4	Pour ouvrir la discussion.....	377
3	Discussion des résultats	377
3.1	Un modèle théorique en mouvement	377
3.2	De quelques réflexions sur la formation au sein du dispositif.....	382
3.3	De quelques réflexions sur la recherche au sein du dispositif	384
4	Les limites de notre recherche	387
4.1	Nous pensions traiter de validité en évaluation, mais... ..	387
4.2	Nous pensions thématiser un jugement professionnel holistique, mais... ..	388
4.3	Au-delà de l'alignement curriculaire élargi, point de salut ?	388
4.4	De quelques biais méthodologiques	389
4.5	Que sont devenus les enseignants de la recherche ?	391
5	Quelques perspectives pour des recherches futures	391
5.1	Considérer la notation comme un objet de recherche à part entière	391
5.2	Partir des pratiques réelles et contextualisées.....	393
5.3	Penser le lien entre recherches collaboratives et communautés de pratiques.....	394
6	Quelques perspectives pour la formation en évaluation sommative notée	395
6.1	Construire la formation continue en référence aux « questions vives de la pratique »	395
6.2	Concevoir la formation initiale autour de connaissances clés	396
7	Une recherche, <i>in fine</i> , pour le développement professionnel de tous les acteurs ?	396
Partie 10 - Bibliographie		398

Table des figures

Figure 1 : Modélisation de l'alignement curriculaire élargi.....	53
Figure 2 : Principe de l'alignement curriculaire selon Gauthier et al. (2005, p. 26)	86
Figure 3 : L'alignement curriculaire selon Anderson (2002, p. 256)	87
Figure 4 : L'alignement curriculaire élargi inspiré d'Anderson (2002) (Pasquini, 2016).....	91
Figure 5: Le processus de décisions des enseignants au travers des pratiques d'évaluation et de notation selon McMillan et Nash (2000, p. 8).....	108
Figure 6 : La modélisation de notre recherche collaborative selon Desgagné et al. (2001) ..	157
Figure 7 : Fernand, épreuve 1, géométrie, page de garde.....	192
Figure 8 : Fernand, épreuve 1, géométrie, tâche 1	193
Figure 9 : Fernand, épreuve 1, géométrie, tâche 2.....	194
Figure 10 : Fernand, épreuve 1, géométrie, tâche 3.....	194
Figure 11 : Fernand, épreuve 1, géométrie, tâche 4.....	195
Figure 12 : Fernand, épreuve 1, géométrie, table de spécification 1	196
Figure 13 : Fernand, épreuve 1, géométrie, table de spécification 2	197
Figure 14 : Fernand, épreuve 2, géométrie, page de garde	198
Figure 15 : Fernand, épreuve 2, géométrie, tâche 3.....	198
Figure 16 : Fernand, épreuve 2, géométrie, tâche 8.....	199
Figure 17 : Fernand, épreuve 2, géométrie, tâche 9.....	199
Figure 18 : Fernand, épreuve 2, géométrie, table de spécification	200
Figure 19 : Paul, épreuve 1, fonctions et algèbre, tâche 1	222

Figure 20 : Paul, épreuve 1, fonctions et algèbre, tâche 2.....	222
Figure 21 : Paul, épreuve 1, fonctions et algèbre, tâche 3.....	223
Figure 22 : Paul, épreuve 1, fonctions et algèbre, tâche 4.....	223
Figure 23 : Paul, épreuve 1, fonctions et algèbre, tâche 5.....	224
Figure 24 : Paul, épreuve 2, calcul littéral, document introductif.....	225
Figure 25 : Paul, épreuve 2, calcul littéral, pondération et notation	226
Figure 26 : Paul, épreuve 2, calcul littéral, pondération et notation (tableau)	227
Figure 27 : Paul, épreuve 2, calcul littéral, page de garde.....	228
Figure 28 : Philémon, épreuve 1, unités composées, page de garde.....	249
Figure 29 : Philémon, épreuve 1, unités composées, tâche 1	250
Figure 30 : Philémon, épreuve 1, unités composées, tâche 2	251
Figure 31 : Philémon, épreuve 1, unités composées, tâche 3a	252
Figure 32 : Philémon, épreuve 1, unités composées, tâche 3b.....	252
Figure 33 : Philémon, épreuve 1, unités composées, tâche 3c	252
Figure 34 : Philémon, épreuve 1, unités composées, tâche 3d.....	252
Figure 35 : Philémon, épreuve 1, unités composées, tâche 4a	253
Figure 36 : Philémon, épreuve 1, unités composées, tâche 4b.....	253
Figure 37 : Philémon, épreuve 2, symétries et homothéties, page de garde	255
Figure 38 : Philémon, épreuve 2, symétries et homothéties, tâche 1.....	256
Figure 39 : Philémon, épreuve 2, symétries et homothéties, tâche 2.....	257

Figure 40 : Philémon, épreuve 2, symétries et homothéties, tâche 3.....	258
Figure 41 : Philémon, épreuve 2, symétries et homothéties, tâche 4.....	258
Figure 42 : Anouck et Alizée, épreuve 1, compréhension de texte, page de garde 1 (objectifs reformulés).....	280
Figure 43 : Anouck et Alizée, épreuve 1, compréhension de texte, page de garde 2 (objectifs du PER et échelle)	281
Figure 44 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 1.....	282
Figure 45 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 2.....	283
Figure 46 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 3.....	284
Figure 47 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), page de garde	286
Figure 48 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), consignes de production	287
Figure 49 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), pondération et notation du TA de fonctionnement de la langue	288
Figure 50 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), échelle de notation	289
Figure 51 : Samia, épreuve 1, expression orale au travers du texte théâtral, items 1 à 5	317
Figure 52 : Samia, épreuve 1, expression orale au travers du texte théâtral, items 6 à 8	318
Figure 53 : Samia, épreuve 1, expression orale au travers du texte théâtral, progressions des apprentissages du PER sélectionnées pour l'objectif L1-34.....	319

Figure 54 : Samia, épreuve 1, expression orale au travers du texte théâtral, feuille de route pour l'évaluation de l'objectif L1-34	320
Figure 55 : Samia, épreuve 1, expression orale au travers du texte théâtral, grille d'évaluation	321
Figure 56 : Samia, épreuve 2, conjugaison, progressions des apprentissages du PER sélectionnées pour l'objectif L1-36.....	323
Figure 57 : Samia, épreuve 2, conjugaison, exercice 1	324
Figure 58 : Samia, épreuve 2, conjugaison, exercice 2	325
Figure 59 : Samia, épreuve 2, conjugaison, exercice 3	326
Figure 60 : Samia, épreuve 2, conjugaison, exercice 4	327
Figure 61 : Samia, épreuve 2, conjugaison, exercice 5	328
Figure 62 : Samia, épreuve 2, conjugaison, échelle.....	328
Figure 63 : Fernand, épreuve 2, géométrie, table de spécification	359
Figure 64 : Paul, épreuve 2, calcul littéral, pondération et notation (tableau)	360
Figure 65 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), échelle de notation	362
Figure 66 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), pondération et notation du TA de fonctionnement de la langue	364
Figure 67 : Modélisation systémique de l'alignement curriculaire élargi.....	378

Table des tableaux

Tableau 1 : Differences between typical tests and authentic tasks (Wiggins, 1988) cité par Mottier Lopez (2015, p. 77).....	99
Tableau 2: Le processus de notation holistique selon Biggs et Tang (2011, p. 7).....	111
Tableau 3: Le dispositif de recherche-formation, déroulement chronologique	144
Tableau 4 : La grille d'analyse croisant les cinq catégories conceptualisantes et les six compétences.....	185
Tableau 5 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Fernand durant la formation.....	202
Tableau 6 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Fernand à la fin de la formation.....	207
Tableau 7 : Indices de développement potentiel des compétences évaluatives de Fernand ..	213
Tableau 8 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Paul durant la formation.....	230
Tableau 9 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Paul à la fin de la formation.....	234
Tableau 10 : Indices de développement potentiel des compétences évaluatives de Paul	240
Tableau 11 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Philémon durant la formation.....	259
Tableau 12 : Eléments disponibles pour l'analyse des points de cohérence et de rupture	

d'alignement curriculaire dans les pratiques évaluatives de Philémon à la fin de la formation.....	263
Tableau 13 : Indices de développement potentiel des compétences évaluatives de Philémon	267
Tableau 14 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Anouck durant la formation.....	294
Tableau 15 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Alizée durant la formation.....	296
Tableau 16 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Anouck à la fin de la formation.....	301
Tableau 17 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Alizée à la fin de la formation.....	304
Tableau 18 : Indices de développement potentiel des compétences évaluatives d'Anouck ..	308
Tableau 19 : Indices de développement potentiel des compétences évaluatives d'Alizée.....	309
Tableau 20 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Samia durant la formation.....	330
Tableau 21 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Samia à la fin de la formation.....	334
Tableau 22 : Indices de développement potentiel des compétences évaluatives de Samia ...	338
Tableau 23 : Phénomènes, questions et réflexions mis en évidence par les enseignants sur	

leurs pratiques de notation avant la formation	347
Tableau 24 : Types de phénomènes, questions et réflexions mis en évidence par les enseignants sur leurs pratiques de notation avant la formation	350
Tableau 25 : Par discipline, phénomènes, questions et réflexions mis en évidence par les enseignants sur leurs pratiques de notation avant la formation	351
Tableau 26 : Mouvements opérés dans les pratiques de notation et nature des obstacles persistants en fin de formation	353
Tableau 27 : mise en perspective des objets de savoir évalués dans la dernière épreuve avec les systèmes de pondération, la notation et les obstacles rencontrés dans les pratiques	357

Partie 1 - Introduction

1 Prolégomènes

Cela fait près de vingt ans que nous sommes aux prises avec des questions, problèmes, dilemmes que les enseignants¹ rencontrent et partagent dans leurs pratiques évaluatives sommatives². Nos motivations à mener cette présente recherche doctorale sont nées d'un parcours premier d'enseignant primaire et secondaire, puis d'un second de formateur. Nous avons éprouvé le besoin de questionner les observations, constats et actions réalisés durant de nombreuses années. Nous avons souhaité entrer dans la « boîte noire » de l'enseignant évaluateur pour appréhender l'origine et le sens de ses pratiques saisies dans des contextes toujours singuliers, souvent stables, mais aussi parfois fragiles ou en questionnement. Il nous a donc semblé important, dans cette introduction, de rendre compte de cet état des lieux et de montrer en quoi il a conditionné notre présent travail. Nous le verrons, le but poursuivi par cette thèse est de questionner la pratique courante et réelle, d'en comprendre mieux les fondements, notamment lorsqu'elle est soumise à l'épreuve d'un modèle théorique dispensé dans une formation. La visée de cette recherche n'est en aucun cas confirmatoire. En effet, nous avons choisi d'approfondir un champ familier pour en saisir la complexité : ce sont parfois les objets les plus proches qui nous échappent le plus.

2 Un parcours parsemé d'interrogations

2.1 Les premières questions de l'enseignant

Notre intérêt pour la problématique de l'évaluation est né de manière spontanée, relativement à nos propres pratiques d'enseignant aux degrés primaires puis secondaires de la scolarité obligatoire, au sein d'équipes de plusieurs disciplines. Au cœur de pratiques très diverses et peu homogènes, confronté à des prescriptions peu stables, pris dans des changements de taille (nouveaux plans d'études et moyens d'enseignement notamment) et des routines, nous avons dû, d'un point de vue d'enseignement et surtout d'évaluation, trouver des manières de faire acceptables à nos yeux tout en affrontant parfois d'importants dilemmes. Durablement inscrit dans une pratique quotidienne alors peu outillée, ce questionnement a ensuite évolué, grâce

¹ Selon la grammaire traditionnelle, le genre masculin n'est pas uniquement l'expression du sexe masculin : il sert aussi de genre commun ou de genre neutre. Ainsi, dans le présent document, les mots de genre masculin appliqués aux personnes désignent les hommes et les femmes.

² Les termes théoriques utilisés dans l'introduction sont définis et explicités dans les deux cadres conceptuels (parties 3 et 4).

notamment à un cursus de formation suivi en Sciences de l'Éducation, en cours d'emploi. Toujours ancré dans la pratique, et inscrit dans un parcours professionnel vécu dans différents établissements vaudois, il s'est progressivement orienté vers des interrogations soucieuses d'une articulation entre pratique et théorie, dans une perspective compréhensive : qu'est-ce qu'une pratique d'évaluation sommative cohérente ? Comment la développer au regard du contexte institutionnel ? Comment favoriser une évaluation sommative qui sert l'élève ? Quels apports permettent des changements dans les pratiques ?

2.2 Les questions du formateur

Dès 2001, nous avons donné forme à ce questionnement dans la conception et la mise en œuvre d'actions de formation continue dispensées aux enseignants de tous les degrés de la scolarité, au sein du Burofco³ d'abord, puis, à partir de 2003, à la Haute École pédagogique du canton de Vaud (HEP)⁴. Là, en tant que formateur, et en équipe, nous avons réalisé une réflexion approfondie à trois niveaux :

- À quelles conditions peut-on optimiser les pratiques évaluatives sommatives en tenant compte du contexte des enseignants ?
- Quels contenus théoriques apparaissent adaptés à ce travail, partant de l'idée qu'il est nécessaire de fournir aux praticiens des connaissances issues de la recherche ?
- Quelles modalités de formation permettent le mieux le développement des compétences évaluatives ?

Ce travail a permis de construire un dispositif de formation continue que nous exploitons encore aujourd'hui, et qui s'inscrit dans le dispositif de recherche mis en œuvre ici, présenté dans la partie 5. C'est dans ce cadre qu'ont été récoltés, lors de multiples interventions auprès de plusieurs centaines d'enseignants, un grand nombre de questions, problèmes, tensions, besoins récurrents, mais non exhaustifs. Issu du terrain, ce matériau a participé en grande partie à délimiter les objets qui ont suscité nos réflexions ici. Il nous a même permis d'affiner et de thématiser ces dernières. Il nous semble donc aujourd'hui pertinent de formuler ces éléments par écrit dans le cadre de cette introduction, parce qu'ils montrent sur quels aspects de leurs

³ Bureau EVM d'organisation de la formation continue. Le Burofco était un organe de formation mis en œuvre par le Département de la formation du canton de Vaud entre 1997 et 2002 pour accompagner la réforme scolaire Ecole Vaudoise en Mutation (EVM).

⁴ Le lecteur se référera à l'annexe Partie 1, 1 pour prendre connaissance des nombreux acronymes mentionnés dans cette thèse.

pratiques des enseignants se questionnent lorsqu'ils se focalisent sur l'évaluation sommative, sur quels points saillants ils demandent des apports, voire des réponses.

Nous sommes conscient que ces éléments initialement récoltés n'ont pas une validité scientifique. Toutefois, empiriquement et de façon relativement informelle, ils constituent des points d'attention qu'il s'agit de rendre visibles, ne serait-ce que pour témoigner de l'influence qu'ils ont eue sur notre travail et dont nous serons amené à discuter. Dans ce sens, nous pensons que cette problématique de la cohérence d'une évaluation sommative porte en elle un enjeu de développement professionnel que l'on peut traduire au travers de questions comme :

- Qu'est-ce qu'une évaluation sommative cohérente, d'un point de vue de forme et de fond ?
- Qu'est-ce qu'un contenu d'épreuve pertinent ?
- Quels contenus évaluer, et quand ?
- Comment comprendre ce qui est à évaluer sommativement dans le plan d'études romand (PER)⁵ ?
- Comment poser des barèmes, construire des échelles de notation, et que demander comme suffisance ?
- Comment gérer au mieux les prescriptions en évaluation dans les pratiques ?

À ce questionnement touchant de multiples dimensions viennent s'ajouter d'autres éléments également observés, voire explicités par les enseignants eux-mêmes dans nos actions de formation continue. Nous avons par exemple relevé très fréquemment qu'une majorité d'entre eux ne se réfère pas au plan d'études pour déterminer les objectifs à évaluer mais plutôt à d'autres référentiels, qu'il existe des habitudes non questionnées dans la manière de construire les épreuves, qu'une tendance à évaluer ce qui a été *fait* par les élèves supplante une évaluation de ce qu'ils ont *appris*, que des perspectives normatives prennent le pas sur des visées plus critériées, que le recours à des outils de notation extérieurs aux apprentissages - comme les échelles élaborées à l'aide d'algorithmes - se généralise sans questionnement de leurs fondements ou idéologies, ou encore, que les modalités d'évaluation au sein des équipes se rigidifient (travaux communs, dates communes fixées à l'avance). La grande majorité des enseignants rencontrés appliquent les directives imposées plutôt que de les adapter à leur contexte. Parfois, certaines de ces directives sont reconstruites au sein de leur établissement par

⁵ Introduit en 2012 dans un vaste mouvement de réforme présenté au point 2.3.1.

les équipes ou la hiérarchie dans un processus de réinterprétation équivoque. Il arrive même occasionnellement de rencontrer des écoles dans lesquelles la direction crée de nouvelles prescriptions, comme le recours aux moyennes cibles ou aux planifications des semaines d'évaluation, sans en expliciter toujours les raisons. Dans ce contexte, nous observons que le recours, voire le retour à des pratiques d'évaluation sommative traditionnelles semble bel et bien réel (Mottier Lopez, 2013, 2014). À ce titre, nous observons ici qu'une attitude passive, soumise à un « on » indéfinissable, est régulièrement présente : « on nous a dit qu'il faut évaluer comme ça », « on prône des seuils de suffisance de tel type », ou « on fait comme le cadre légal nous dit de faire », sont des propos courants que les participants à notre recherche ont également produits. Comment comprendre ce rapport à la prescription ? C'est un des axes que nous aborderons dans ce travail.

Malgré cela, la majorité des enseignants ne sollicite pas de formation sur cet objet. Et l'on pourrait nous rétorquer que, contrairement à nos observations, ce silence est signe de stabilité, preuve que les pratiques sont en place et donnent majoritairement satisfaction. C'est une explication parmi d'autres. Nous privilégierons toutefois les deux hypothèses interprétatives suivantes, comme autant de guides pour notre recherche :

- Pris dans la mouvance d'une école en mutation et chargés par des tâches de plus en plus administratives, bon nombre de praticiens mettent l'accent sur des urgences, et l'évaluation sommative n'en constitue pas une. En effet, beaucoup d'équipes trouvent leur compte dans des arrangements qui satisfont tous les partenaires concernés, notamment dans une forme d'évaluation menée en collectif qui rassure. Là, ils peuvent assumer certains compromis, revendiquer une mutualisation des exigences et des pratiques comme gage de justice, et donc vivre avec des doutes et des dilemmes qui les remettent éventuellement en question, mais pas au point de se former ;
- Il existe des enseignants qui adhèrent aux directives imposées parce qu'elles confortent leurs valeurs, leurs conceptions de l'enseignement, de l'apprentissage, de l'évaluation. Certains défendent explicitement des logiques méritocratiques qu'ils disent même retrouver dans les textes de loi. Dès lors, nous faisons l'hypothèse que les pratiques évaluatives renvoient à des dimensions de posture et d'identité professionnelle profondément ancrées.

De surcroît, nos observations nous ont amené à penser que le développement des compétences évaluatives dépend de facteurs contextuels. Nous en voudrions pour preuve la difficulté à

recruter huit enseignants⁶ pour notre recherche (voir l'historique à la partie 5). Conjointement, l'évaluation sommative est une trace produite par l'élève qui sort de la classe. Cette dimension sociale influence la perception qu'en ont les enseignants. On comprend alors qu'au-delà d'une question de compétences ou d'intérêt, évaluer et noter⁷ est parfois décrit comme douloureux, voire pénible, risqué, et surtout, chronophage. Certains enseignants, au travers de leur pratique évaluative, disent même ressentir un effet d'évaluation de leur propre enseignement. Mais l'évaluation reste incontournable, nécessaire, puisqu'il ne s'agit ni plus ni moins de rendre des comptes, de faire preuve de transparence, de valider, en d'autres termes, de remplir également un devoir institutionnel, ritualisé, normé, légiféré. Une tension existe donc à ce niveau, que nous traiterons également, mais qui nous permet déjà ici de formuler quatre hypothèses au service de la définition de notre problématique de recherche.

2.2.1 Hypothèse 1 : la recherche d'une évaluation sommative « juste » est au cœur des préoccupations des enseignants et constitue donc une entrée pertinente pour travailler la cohérence des pratiques évaluatives

C'est la demande la plus récurrente en formation continue, et qui est ressortie chez une majorité des enseignants de notre recherche. Mais que traduit cet adjectif ? Donner des verdicts sur les apprentissages des élèves correspondant au mieux à ce que ces derniers seraient censés maîtriser ? Fournir des renseignements sur les acquis, mais aussi sur les lacunes des élèves, dans une perspective de régulation ? Offrir une meilleure compréhension du verdict, les résultats devant être facilement communicables et recevables par les élèves, les parents et les autres acteurs de l'école ? Ou une conjugaison de plusieurs de ces sens ? Nous tenterons d'y répondre.

2.2.2 Hypothèse 2 : les disciplines scolaires jouent un rôle fondamental dans la formulation des réflexions et questions des enseignants

Du point de vue des pratiques observées et des demandes des participants à la recherche, nous observons qu'il existe des habitudes évaluatives spécifiques aux disciplines. Certaines semblent

⁶ Nous ne considérerons finalement que six d'entre eux, pour des raisons méthodologiques que nous expliquons dans la partie 5.

⁷ D'un point de vue de pratique, nous faisons un rapport direct entre évaluation sommative et notation, puisque dans le canton de Vaud, toute évaluation de ce type aboutit à une note. Toutefois, d'un point de vue théorique, nous différencierons pratique d'évaluation sommative et pratique de notation (voir le cadre conceptuel en partie 3).

émaner de traditions didactiques ou d'habitudes héritées des propres parcours scolaires des enseignants (évaluer le vocabulaire en langues, les livrets en mathématiques, pour ne citer que celles-ci), d'autres relever de la compréhension des manuels à disposition (évaluer des problèmes mathématiques identiques à ceux exercés en classe, se fier aux modules en rapport avec une démarche didactique donnée dans une séquence « officielle » de français), d'autres encore participer de certains rituels d'équipe ou d'établissement (évaluer le théâtre dans le cadre de la fête de Noël, les jeux mathématiques réalisés pour la préparation d'un concours, ou les multiples objectifs travaillés dans le cadre d'un projet interdisciplinaire de fin de scolarité). Parallèlement à ces constats, nous interrogerons le statut des tâches emblématiques d'évaluation qui paraît exister dans les pratiques : la production écrite en français (qui prend différentes formes en fonction de l'âge et du niveau des élèves) évaluée par des critères centrés également sur le fonctionnement de la langue, et la résolution de plusieurs problèmes de complexité croissante au sein d'une même épreuve en mathématiques, pour ne citer que ceux-ci. Cela constituera une partie conséquente de nos analyses.

2.2.3 Hypothèse 3 : la construction d'épreuves et la pondération sont des pratiques relativement peu outillées

Parallèlement à ces spécificités, nous observons que bon nombre de réflexions et de questions d'enseignants se rejoignent. Elles portent sur deux points majeurs, et se sont retrouvées chez les enseignants participant à la recherche : la construction de l'évaluation et la pondération des performances attendues. Le premier point renvoie à l'agencement des tâches et à leur rapport aux apprentissages évalués. Là, nous avons remarqué que les enseignants sont plutôt dans une logique ascendante, c'est-à-dire qui part des tâches puis vers leur rapport avec des objectifs. À ce niveau, il n'est pas rare de constater que ce rapport n'est pour une bonne partie d'entre eux pas simple à établir. Si produire ou choisir des tâches d'évaluation est une pratique qui semble relativement routinière, dire quels apprentissages ces tâches évaluent, et donc, quels objets disciplinaires y circulent, est bien plus complexe. L'analyse de ces rapports sera au cœur de notre problématique.

Le second point, relatif aux modalités de pondération des performances, renvoie à l'établissement des seuils de suffisance et, corollairement, à la construction des échelles. C'est notamment à ce niveau que nous rencontrons souvent un important questionnement chez les enseignants, toutes disciplines confondues, au niveau des choix qu'ils réalisent et dans leurs façons de les expliciter ou de les justifier. C'est également ici que les participants à la recherche

sont massivement intervenus. Le doute principal réside fréquemment dans l'établissement des seuils de suffisance, qui renvoie lui-même à la question du minimum exigé, et donc aux éléments disciplinaires fondamentaux à maîtriser. Ce doute est par ailleurs renforcé dans certains degrés du secondaire, les directives régissant les modalités de promotions entre filières étant uniquement basées sur des moyennes. L'enjeu pédagogique sur l'atteinte de la note suffisante est ici très élevé. Un raisonnement pédagogique serait donc possible ici, nous en parlerons abondamment d'un point de vue conceptuel et empirique.

Ceci étant, durant ces nombreuses années, nous avons remarqué que la pratique de pondération qui domine, toutes disciplines confondues, est la pose de points, de demi-points, voire de quarts de points. Le recours aux critères qualitatifs est irrégulier, et spécifique à certaines disciplines, voire à quelques tâches d'évaluation emblématiques, comme la production de texte en langues. Par ailleurs, lorsqu'une approche qualitative est engagée, elle se retrouve, *in fine*, rattrapée par une réduction des éléments qualitatifs des apprentissages en une somme de points, encore. Les enseignants se demandent donc comment élaborer des évaluations qualitatives, tout en mettant en évidence cette tension avec l'obligation de construire une note. Une tension, toutefois, fort compréhensible. En effet, jusqu'où une approche qualitative est-elle possible en évaluation sommative lorsque prédomine le recours systématique aux totaux de points, aux pourcents et aux moyennes, dans les pratiques et les directives, plus ou moins explicitement ? Nous aborderons spécifiquement ces questions d'un point de vue théorique dans la partie 3, dans l'analyse de nos données dans les parties 8 et 9, puis dans notre conclusion.

2.2.4 Hypothèse 4 : le questionnement de départ des enseignants se réfère surtout à la notation parce qu'elle représente la part visible de toute évaluation

Nous constatons qu'une grande partie des problèmes évoqués par les enseignants se réfère à la notation⁸. Il en a été de même dans le cadre de notre étude. À ce niveau, les enseignants semblent être en insécurité face à ce qu'ils sont en droit de faire, ou de ne pas faire. S'ils sont très peu outillés pour passer de l'élaboration de l'évaluation à la pondération des apprentissages évalués, la phase de construction de la note est toute aussi délicate, si ce n'est carrément problématique. Beaucoup de questions tournent spécifiquement autour de la réalisation des échelles : « peut-on les créer ou doit-on exploiter uniquement les logiciels mis à disposition par

⁸ Dans le canton de Vaud, les enseignants sont amenés à noter sur une échelle de 1 à 6, avec demi-notes, ce qui représente onze échelons de notation. Le seuil de suffisance est la note 4.

le Département ou disponibles sur la toile ? » ; « sommes-nous obligés de les utiliser tous de la même manière ? », « quand doivent-elles être élaborées ? ». Viennent ensuite les doutes relatifs à la valeur des notes : « qu'est-ce qu'un 4 ? », « suis-je astreint à utiliser tous les échelons disponibles ? », « je ne souhaite pas mettre de 1, en ai-je le droit ? », « quelles différences existe-t-il entre un 5 et un 5.5 ? ». Autant de questions que nos enseignants ont également formulées.

En revanche, et malgré les spécificités relevées dans notre deuxième hypothèse, les questions relatives aux particularités disciplinaires et à leur impact sur les démarches d'évaluation sont plus rares. À ce niveau, nous constatons que les enseignants font preuve d'une forme d'assurance, s'appuyant sur leur connaissance des programmes ou sur leur expérience. Ce bagage leur permettrait de gérer ces spécificités. Là, nous faisons l'hypothèse suivante : le fait qu'ils échangent ou se réfèrent massivement à du matériel existant (fiches, dossiers, manuels, outils trouvés sur Internet ou dans des banques de données diverses) est une des causes principales de cette absence de questionnement. Dans nos actions de formation continue, c'est en traitant de la cohérence dans l'évaluation sommative que nous arrivons aux questions de contenus dans l'enseignement et l'apprentissage, même si, comme évoqué, ce n'est pas à ce niveau que leurs interrogations se situent. Dans le cadre de notre recherche, un questionnement relatif aux caractéristiques des objets de savoir dans les évaluations a régulièrement été soumis aux participants. Nous détaillons son impact dans nos chapitres de résultats.

Tournons-nous maintenant plus finement vers le questionnement du chercheur.

2.3 Des questions pour le chercheur

Cette thèse vise à formaliser ce questionnement, en passant d'un questionnement professionnel ou de formation à un questionnement de recherche.

2.3.1 Quelques éléments contextuels pour saisir l'environnement de la recherche

Le contexte dans lequel s'inscrit cette recherche est particulier. L'école vaudoise traversait une période de changements sans précédents lorsque notre dispositif a été mis en œuvre avec les enseignants en 2014-2015. Depuis 2011, un nouveau plan d'études romand (PER) a été progressivement introduit. En 2013, la Loi sur l'enseignement obligatoire (LEO) a remplacé

l'ancienne loi scolaire de 1984, en application du Concordat HarmoS⁹. La structure de l'école obligatoire a évolué. Les grilles horaires ont été modifiées. De nouvelles directives relatives à l'évaluation ont été formulées, légiférant plus strictement les modalités de certification et de promotion des élèves d'une année à l'autre, voire d'un cycle à l'autre, surtout dans les années 9-11 qui clôturent la scolarité obligatoire.

Les degrés secondaires, au nombre de trois, ont été très touchés par ces changements. En sus de ceux évoqués plus haut, nous noterons encore l'arrivée progressive de nouveaux moyens d'enseignement (manuels, guides du maître, pistes didactiques), le passage de trois voies (VSO, VSG et VSB) à deux (VG et VP), la Voie Générale étant subdivisée en deux niveaux 1 (exigences de base) et 2 (exigences supérieures), en fonction des compétences des élèves. Des groupes de disciplines ont été créés afin de définir les seuils de promotion d'un niveau à l'autre et d'un cycle à l'autre¹⁰. Des passerelles entre cycles et niveaux se sont multipliées, rendant le processus d'orientation des élèves toujours possible au semestre en 9e et 10e années. Le recours aux notes, aux moyennes de disciplines, aux moyennes de groupes de disciplines s'est généralisé et est devenu la référence pour l'établissement des pronostics des élèves¹¹.

Nous faisons l'hypothèse que ce contexte en changement a un impact fort sur les pratiques d'évaluation sommative des enseignants. Aux prises avec des prescriptions au nombre rarement égalé, confrontés à de nouvelles références influençant directement leur quotidien, leurs pratiques d'évaluation sommative peuvent-elles être innovantes, et dans l'affirmative, à quelles conditions ? Paradoxalement, peut-on envisager que c'est en période de changement que les habitudes peuvent aussi vaciller et provoquer un questionnement propice à la formation, pour autant qu'un cadre approprié soit fourni ? Nous abordons ces questions dans ce manuscrit.

On le comprend, il nous semble fondamental que la recherche en évaluation approfondisse encore davantage la problématique de l'évaluation sommative, sachant que sa compréhension est indispensable au développement d'une école juste et exigeante. Il devient à nos yeux important d'investiguer et de comprendre à quelles conditions les compétences évaluatives des enseignants gagnent en cohérence, en référence à leurs disciplines, leurs contextes, leurs publics, mais aussi par rapport à leurs contraintes ou leurs croyances, afin de mieux appréhender

⁹ <http://www.edk.ch/dyn/11737.php>.

¹⁰ http://www.vd.ch/fileadmin/user_upload/organisation/dfj/dgeo/fichiers_pdf/Flyer_Cycle3_9-10_2015.pdf.

¹¹ Pour visualiser ces systèmes et leurs spécificités, le lecteur se référera à l'annexe Partie 2.

comment ils jugent leurs élèves d'un point de vue sommatif.

3 Les enjeux relevés par la thèse

Les enjeux de la thèse que nous présentons ici sont donc à nos yeux de trois ordres au moins. Le premier est relatif au développement de connaissances pour la pratique. Là, il s'agira d'apporter des éléments sur les pratiques d'évaluation sommative réelles et sur le développement des compétences évaluatives des enseignants. Comment ces derniers s'y prennent-ils, concrètement, pour évaluer leurs élèves d'un point de vue sommatif ? Sur quels aspects de cette pratique leur attention porte-t-elle prioritairement et pourquoi ? Comment exploitent-ils le plan d'études romand, ou d'autres références ? Comment font-ils pour élaborer leurs épreuves au sein de leur discipline ? Quels processus mettent-ils en œuvre pour noter les travaux de leurs élèves ? Quelles connaissances leur permettent de progresser dans leurs contextes respectifs, sans sacrifier leurs habitudes ni occulter leurs obligations ?

Le second enjeu se rapporte à la production de savoirs pour la recherche. Comment comprendre l'interaction d'un modèle théorique avec les pratiques évaluatives sommatives réelles, ce modèle étant l'alignement curriculaire élargi que nous présenterons et argumenterons ci-après ? En quoi apporte-t-il des pistes pour développer des savoirs pour la recherche ? Dans quelles mesures les pratiques peuvent-elles en questionner la pertinence ? Quelles influences une modalité de recherche de type collaboratif (Desgagné, 2001) a-t-elle sur la production des données et leur analyse ? Quels apports et/ou limites la perspective théorique de l'*assessment literacy* offre-t-elle pour appréhender le développement de compétences évaluatives chez les enseignants ?

Ces interrogations, non exhaustives, devront se conjuguer à notre troisième et dernier enjeu : une nécessaire mise à distance de nos questions spontanées d'enseignant et de formateur, que nous devrons discuter à la lumière de nos choix épistémologiques et théoriques ainsi que des résultats obtenus.

4 Organisation du manuscrit

Le manuscrit s'organise de la manière suivante :

- La partie 2 présente une large revue de la littérature qui porte sur divers champs. Elle commence par exposer les résultats de recherches significatives sur les pratiques

évaluatives sommatives puis sur la notation, en mettant en évidence le manque de problématisation existant au niveau de leur possible articulation. Elle présente ensuite les travaux qui ont abordé le modèle théorique exploité ici, l'alignement curriculaire, et met en évidence la nécessité de le conceptualiser de manière « élargie ». Conformément à cette acception, elle interroge les apports sur la pondération puis sur la notation lorsque ces deux composantes sont saisies dans une recherche de cohérence curriculaire. Ces apports faisant tous plus ou moins explicitement référence aux questions de jugement professionnel, de validité et de paradigmes en évaluation, elle se termine en les mettant en perspective.

- La partie 3 expose le cadre conceptuel lié aux pratiques évaluatives sommatives. Elle théorise l'élargissement du modèle de l'alignement curriculaire du point de vue de l'évaluation sommative notée. Y figurent par ailleurs les choix conceptuels adoptés en référence aux composantes du modèle, à savoir la référenciation restreinte, le *design*, la pondération et la notation. Les concepts de validité et de jugement professionnels sont également abordés, puisqu'ils interviennent en partie dans la compréhension des pratiques.
- La partie 4 est également théorique : elle s'attache à définir et à conceptualiser la perspective théorique de l'*assessment literacy*, ou littératie en évaluation, que nous exploitons comme cadre interprétatif du développement des connaissances et compétences des enseignants.
- La partie 5 concerne la recherche entreprise d'un point de vue contextuel. Nous y expliquons notre stratégie de recrutement des enseignants, le profil des candidats retenus, ainsi que les caractéristiques du dispositif de recherche collaborative mis en œuvre.
- La partie 6 présente les orientations méthodologiques choisies. Nous y exposons les raisons qui nous ont fait privilégier une méthode de recherche collaborative, ainsi que la visée interprétative et compréhensive de la recherche. Vient ensuite une partie présentant les données récoltées et leurs modes de structuration et d'analyse, notamment au travers de catégories conceptualisantes et de processus de triangulation.
- La partie 7 rend compte de nos premiers résultats. Elle expose, interprète et discute les analyses de nos données. Elle traite nos deux premières questions de recherche. Structurée au travers de six portraits (un par enseignant), elle donne à voir leur évolution tout au long du dispositif de recherche, au niveau des mouvements dans leurs pratiques évaluatives et du développement potentiel de leurs compétences.

- La partie 8 expose, interprète et discute les analyses de nos données relatives à notre troisième question de recherche. Là, en termes de résultats, c'est à une analyse transversale que nous nous livrons (c'est-à-dire entre les enseignants), en nous focalisant sur le développement potentiel de leurs pratiques de notation.
- La partie 9 clôt cette thèse par une conclusion faisant état de nos principaux résultats, mais proposant également une discussion de ceux-ci et de notre modèle théorique. Les limites de notre recherche et les perspectives qu'elle ouvre modestement sont enfin abordées à la lumière de références complémentaires.

Partie 2 - Revue de littérature

1 Introduction

Dans cette partie, au travers des multiples sources consultées, nous souhaitons montrer la portée de certains travaux sur notre thématique, mais également que certains objets de recherche en rapport avec les pratiques effectives d'évaluation sommative demandent encore à être investigués. Pour commencer, cette revue délivrera donc des apports de la recherche qui documentent l'état actuel des connaissances sur les pratiques évaluatives sommatives des enseignants, afin de poser les premiers jalons de notre problématique. Nous verrons dans un premier temps qu'il émerge de nombreux travaux des constats mettant en évidence nombre de tensions :

- Concernant la perception que les enseignants ont de leurs compétences évaluatives et de ce qu'est pour eux une évaluation sommative cohérente, au regard de la réalité de leur pratique ;
- Relativement au statut des objets de savoir dans ces pratiques ;
- Au niveau des pratiques de pondération et de notation, dont la littérature souligne la spécificité.

Suivra une partie conséquente dans laquelle nous aborderons la littérature relative au modèle théorique de notre recherche, en nous référant aux travaux qui traitent du concept d'alignement curriculaire ainsi que des composantes qui le caractérisent : le curriculum, le référentiel, les objectifs et les objets, les tâches évaluatives et l'évaluation. Tous ces éléments seront dès lors présentés à l'aune de la cohérence de l'alignement curriculaire, que nous expliciterons également en nous appuyant sur des recherches ciblées autour de la notion de taxonomie des habiletés cognitives. Puis, dans une perspective systémique et dynamique, nous pointerons les apports saillants que l'alignement curriculaire développe en matière de compréhension de la pratique évaluative sommative, mais aussi les limites de sa conceptualisation de départ, afin d'aboutir à une acception « élargie » jusqu'à la notation. Pour cela, nous ferons état des travaux qui ont étudié les pratiques relatives aux processus de construction de la note au sein du modèle.

Nous proposerons ensuite une section sur les recherches significatives qui ont abordé la problématique du jugement professionnel en évaluation, en faisant l'hypothèse qu'il est omniprésent dans notre modèle théorique, et qu'une de ses finalités est d'aboutir à une évaluation dont la validité est optimale. Cette question de validité étant omniprésente dans les

sources consultées, nous en présenterons dès lors différentes acceptions. Pour cela, nous convoquerons d'un point de vue critique des études qui énoncent ses multiples définitions, et qui nous permettront de justifier nos choix.

Étant donné que la validité est souvent mise en perspective avec les paradigmes qui ordonnent les cadres compréhensifs de la pratique évaluative, nous poursuivrons cette revue en présentant les travaux qui montrent la complexité inhérente à leurs définitions, mais surtout l'impact qu'ils ont sur le développement des outils d'évaluation. Nous terminerons en explicitant notre orientation conceptuelle.

2 L'étude des pratiques d'évaluation sommative, ou l'expression de tensions

Que savons-nous des pratiques d'évaluation sommative des enseignants ? Quels phénomènes mettent-elles en évidence ? Quels problèmes y rencontre-t-on ? Une importante production scientifique dans le champ de l'évaluation a abordé ces questions et fournit des tentatives de réponses. Nous commencerons donc par mentionner ses apports, car ils ont participé à mieux circonscrire notre objet de recherche, mais également ses limites. Nous nous référerons ensuite à des travaux significatifs du champ de la didactique pour mettre en perspective ou enrichir nos axes de problématisation. Mais un constat réunit déjà la quasi-totalité des recherches parcourues : leurs résultats sont interpellants, tant au niveau de l'étude des compétences évaluatives des enseignants et de leur connaissance en la matière, que des pratiques énoncées.

De manière à concentrer notre recension, nous nous sommes attelé principalement à l'étude de travaux récents, réalisés depuis 2000. Nous assumons ce parti pris. En procédant de la sorte, nous avons souhaité nous baser sur des recherches actualisées, un bon nombre d'entre elles reprenant des résultats d'études antérieures en les éprouvant dans des contextes nouveaux¹². Cela ne signifie en rien que les apports réalisés avant ce terme n'ont aucune valeur à nos yeux, bien au contraire : ils ont été la source de ceux qui nous ont aidé à problématiser notre objet.

Enfin, on pourra critiquer ces recherches en pointant le fait qu'elles sont massivement des enquêtes sur des pratiques déclarées (survey) ou aux méthodologies quantitatives, et que dès lors, elles ne rendent pas compte d'un agir en situation, possiblement préhensible tant au niveau des processus évaluatifs en jeu que des produits réalisés. Cela nous semble en partie justifié. Il

¹² Par exemple, nous sommes conscient que les recherches sur l'alignement curriculaire ont été fortement inspirées et se sont développées suite aux travaux sur les écoles efficaces qu'Edmonds a menés aux Etats-Unis dès 1979.

est toutefois important de préciser ici que nous avons peiné à trouver des études scientifiques sur les pratiques évaluatives réelles, car elles ont suscité peu d'états des lieux des pratiques des enseignants dans leurs classes (CNESCO, 2014).

2.1 L'évaluation normative : une évaluation par défaut

Martinez, Stecher et Borko (2009) ont étudié les pratiques évaluatives sommatives d'enseignants du primaire aux États-Unis en langues et en mathématiques¹³. Ils montrent que ces derniers s'inscrivent souvent dans une perspective d'évaluation normative basée sur l'hypothèse d'une distribution gaussienne des performances des élèves, sans en connaître les enjeux et incidences. Goldberg et Roswell (2000) font le même constat sur la production écrite. Aux États-Unis encore, ces résultats sont confirmés par Knight et Yorke (2003), deux chercheurs ayant travaillé à référencer la prévalence des pratiques d'évaluation sommative. Ils observent que l'évaluation normative est, tous degrés confondus, l'évaluation par défaut.

Ce constat est appuyé également du côté francophone avec Antiby (2003) et ses recherches sur la constante macabre - nom donné à la proportion indéfectible d'élèves en échec lorsqu'une évaluation normative est mise en œuvre. Il montre que 96% des enseignants sondés reconnaissent cette constance et que 89% souhaitent s'en affranchir. Il serait toutefois délicat de généraliser ces résultats. En effet, même si la perspective normative ainsi définie paraît très courante dans les pratiques d'évaluation sommative, d'autres chercheurs, dans des contextes différents et à l'aide de méthodologies plus compréhensives, ont montré que les enseignants recourent également et de manière régulière à l'évaluation critériée ou y sont sensibles (e. g., Mottier Lopez & Allal, 2008 ; Mottier Lopez & Laveault, 2008, Mottier Lopez & Pasquini, 2017).

2.2 Des pratiques évaluatives contrastées

Parallèlement, nombre de recherches constatent globalement que les pratiques évaluatives sommatives évoluent très peu. Par exemple, McKinney, Chappell, Berry & Hickmann (2009) mettent en évidence qu'en mathématiques, seul un pourcentage restreint d'enseignants met en œuvre des démarches d'évaluation sommative alternatives centrées sur les apprentissages et les progrès des élèves.

¹³ Nous avons à dessein choisi de documenter essentiellement des recherches inscrites dans ces disciplines, comme cette thèse.

Un constat confirmé au niveau international par Nusche, Radinger, Santiago et Shewbridge (2013) dans une méta-analyse qui, pour nous, fait référence à large échelle. Les auteurs affirment que dans la plupart des pays de l'OCDE les évaluations restent traditionnelles, c'est-à-dire centrées sur un nombre limité d'apprentissages et d'objets de savoir, réduisant ainsi le contenu du curriculum. Mottier Lopez (2014) observe aussi de telles pratiques, qu'elle définit comme étant normées, axées fortement sur les notes et les moyennes, et ceci surtout en période de changement - comme c'est le cas aujourd'hui dans notre système scolaire.

Moss (2013), dans une revue de littérature constituant également pour nous une ressource clé, montre qu'une des raisons de cet immobilisme viendrait d'un manque de formation initiale et continue des enseignants à l'évaluation sommative, pourtant centrale. Elle met en évidence, en se référant à des travaux inscrits dans des contextes variés et réalisés dans des méthodologies parfois mixtes, que les enseignants utilisent régulièrement une variété de méthodes d'évaluation (formative, sommative) alors que leurs connaissances en la matière sont jugées faibles (Mertler, 2004, 2009). Là, l'évaluation sommative serait davantage le fruit d'habitudes peu questionnées ou d'essais-erreurs que de pratiques situées théoriquement. Cela met au centre la question des connaissances à posséder pour évaluer d'un point de vue sommatif¹⁴. Des travaux pointent cette question, et montrent clairement que l'acquisition de connaissances en évaluation a un impact important sur la perception que les enseignants ont de leurs propres compétences évaluatives (Zhang & Burry-Stock, 2003 ; Popham, 2009). Ces auteurs vont même jusqu'à dire que posséder des connaissances en la matière permet, dans une certaine mesure, d'outiller la pratique et la rendre plus valide, et donc, de compenser le manque d'expérience. Il y aurait donc urgence à penser la formation sur cet objet, surtout lorsque ce sont les élèves qui subissent les conséquences négatives de ces pratiques.

Cela dit, d'autres recherches nuancent ces premiers constats. Dans un rapport très informé mené auprès de centaines d'enseignants du collège en France, et en croisant des entretiens avec des questionnaires, Braxmeyer, Guillaume, Lairez et Levy (2005) ont documenté les pratiques d'évaluation sommative en cours. Ils mettent en évidence des éléments très intéressants pour notre problématique. Tout d'abord, une grande majorité d'enseignants disent évaluer essentiellement des compétences et des connaissances et qualifient cette pratique comme très solitaire. Les pratiques exprimées sont décrites comme transparentes et justes, c'est-à-dire

¹⁴ Cette question est abordée conceptuellement dans la partie 4 sur l'assessment literacy, puis dans nos deux chapitres de résultats.

qu'elles font l'objet de communications régulières et explicites tout en visant les progrès du plus grand nombre. Et cela, comme le relèvent les chercheurs, même s'il existe des différences importantes entre les disciplines scientifiques et littéraires. Un consensus émerge cependant autour du fait que la majorité des enseignants donnent les critères d'évaluation et leur pondération aux élèves, ce qui infirme certains autres résultats. Au niveau de la construction des évaluations, les chercheurs montrent que les enseignants utilisent essentiellement les exercices proposés dans les manuels scolaires afin « d'y évaluer principalement la mobilisation de connaissances » (p. 48). Le rapport ne dit toutefois pas de quelles connaissances il s'agit, et comment les enseignants évaluent cette mobilisation. Son contenu décrit davantage des pratiques d'évaluation situées, conscientes, assumées. Ce contrepoint est intéressant pour nous, car il nuance les constats relevés plus haut. En effet, nous faisons ainsi l'hypothèse que les enseignants *font avec* leur expérience, s'arrangent (Merle, 2007) pour évaluer au mieux leurs élèves. Et que ces arrangements font partie intégrante de la pratique en donnant, somme toute, satisfaction.

Toutefois, nous pensons que cette tension entre des recherches aux méthodologies assez semblables confirme une incomplétude, et qu'il s'agit d'aller plus loin en investiguant les pratiques réelles et pas seulement déclarées.

2.3 De l'(in)cohérence et de l'(in)validité

De leur côté, Black, Harrison, Hodgson, Marshall et Serret (2010) ont montré que les pratiques évaluatives sommatives des enseignants ne sont souvent pas en cohérence avec les conceptions que ces derniers ont de la validité¹⁵ de l'évaluation. Là, seul un travail critique en formation, en partant de situations réelles, amène les enseignants à questionner leurs pratiques, voire à les modifier. Dans le même sens, des travaux mettent en évidence le fait qu'il existe de forts écarts entre les représentations que se font les enseignants des pratiques d'évaluation sommative cohérentes et la description de leurs propres pratiques (McKinney & al., 2009 ; Rieg, 2007), notamment lorsque ces dernières sont influencées par le contexte d'établissement (Mertler, 2009), ou par des épreuves externes (McMillan, 2003, 2005).

Dans ce registre, certaines études sur les pratiques réelles sont intéressantes. Comme celle de

¹⁵ La validité est définie par les chercheurs cités comme le degré de congruence existant entre ce que l'on souhaite évaluer et ce qui est réellement évalué. Nous discutons son statut au cœur de notre recherche et dans notre conclusion.

Bateman, Taylor, Janik et Logan (2009), réalisée en CEGEP¹⁶ au Québec, et qui documente qu'à objectifs communs, des enseignants de même discipline n'enseignent pas les mêmes contenus, ni n'évaluent les mêmes apprentissages à la fin de leurs séquences. Ces chercheurs soulèvent ici la complexité à penser une forme de cohérence dans les pratiques évaluatives sommatives, et cela même si les enseignants s'accordent sur un certain nombre d'éléments, comme les objectifs d'apprentissages à poursuivre et évaluer, ou le déroulement des séquences. De fait, ces travaux montrent que toute pratique d'évaluation sommative est située et dépendante d'un contexte.

2.4 Des contextes très influents

Ces constats interrogent tant ils touchent des compétences clés de la pratique enseignante. L'imperméabilité à la nouveauté ou à la réflexivité observée dans les pratiques peut provenir parfois également de contraintes extérieures, de normes prescrites et de réalités de classes parfois difficiles (McMillan & Nash, 2000). Ces éléments externes peuvent brouiller les pratiques que les enseignants, dans ces recherches, décrivent comme intimement personnelles, en les amenant à mal interpréter les résultats des élèves et mésestimer leurs habiletés (Kilday, Kinzie, Mashburn & Whittaker, 2011). Il y a donc urgence à penser à quelles conditions les pratiques évaluatives sommatives peuvent gagner en cohérence de manière générale, mais surtout lorsqu'elles sont aux prises avec des éléments externes qui inévitablement les conditionnent, comme les prescriptions (Mottier Lopez & Pasquini, 2017), les épreuves externes (Squires, 2009), les logiques d'équipe (Lafortune & Turcotte, 2008) ou encore les modes de leadership (Pelage, 2003).

2.5 Les principales limites des recherches convoquées

Un point commun relie ces travaux : les enseignants, tous cycles confondus, apparaissent insuffisamment compétents dans le domaine et semblent manquer de confiance dans leurs pratiques d'évaluation sommative (Black et al., 2010). Comme évoqué plus haut, il manque, selon nous, des résultats de recherche sur les pratiques réelles d'évaluation sommative, à visée compréhensive et interprétative, ce que confirme Moss (2013) lorsqu'elle affirme : « there is a need for research designs that go beyond teachers' self-report, surveys, and inventories » (p.

¹⁶ Au Québec, un CEGEP est un collège d'enseignement général et professionnel, un établissement d'enseignement collégial public où sont offertes des formations techniques et préuniversitaires.

252). Notre problématique et méthodologie de recherche s'inscriront dans cette perspective.

2.6 Des travaux sur l'évaluation sommative issus des didactiques

De nombreuses sources le mettent en évidence : les disciplines impactent les pratiques d'évaluation sommative ou, pour le dire autrement, « il y aurait bien des comportements évaluatifs disciplinaires spécifiques » en évaluation (Braxmeyer et al., 2005, p. 52). Certains travaux évoquent l'influence de la discipline sans la détailler (e.g., Black & al., 2010 ; Zhang & Burry-Stock, 2003). D'autres l'approfondissent, et l'on dispose de nombreuses recherches qui s'inscrivent explicitement dans une discipline. À ce niveau, toutes montrent que les caractéristiques des objets de savoir évalués influencent grandement la pratique évaluative sommative des enseignants (e. g., Biberman-Shalev, Sabbagh, Resh & Kramarski, 2011 ; Klein et al., 2009; Meier, Rich & Cady, 2006). Des habitudes et *des façons de faire* existent donc en évaluation sommative et se différencient en fonction des disciplines scolaires (Braxmeyer et al., 2005).

Mottier Lopez et Laveault (2008) affirmaient déjà il y a une dizaine d'années que les pratiques d'évaluation devraient se diversifier en fonction, notamment, « des choix didactiques et des objectifs qui en découlent » (p. 13). Cela dit, relativement à l'importance des contenus, force est de constater qu'à ce jour, les travaux francophones en didactique traitant de la question sont encore assez rares, comme si l'enseignement et l'évaluation sommative étaient, dans une certaine mesure, dissociés. Un constat appuyé par Millon-Fauré (2013) : « Bien qu'il s'agisse d'un rituel quasi-quotidien de la classe, les évaluations constituent encore des dispositifs peu étudiés en didactique » (p. 149). L'intention est donc présente, mais les résultats encore assez discrets. Le rôle de la didactique dans l'évaluation sommative semble pourtant de plus en plus partagé au sein de la communauté scientifique. Nous disposons en effet depuis peu de travaux qui s'inscrivent clairement dans l'étude des pratiques relativement à des objets disciplinaires circonscrits (e. g., Demonty, Fagnant & Dupont, 2015 ; Sayac, 2018)¹⁷. Dès lors, pour documenter ce que nous apprennent les travaux de didactique en la matière, nous nous concentrerons sur le français langue première et les mathématiques, les deux disciplines que nous avons prises en considération dans notre étude. Commençons par la didactique de la

¹⁷ Voir également l'ouvrage publié par l'Association Internationale pour la Recherche en Didactique du Français (AIRDF) en 2016 « L'évaluation en classe de français, outil didactique et politique », ou les résumés des communications du colloque « Evaluation & mathématiques : dispositifs, validité et pratiques » organisé par l'Ecole Supérieure du Professorat et de l'Education (ESPE) de l'académie de Créteil, composante de l'Université Paris Est Créteil (UPEC) en 2016.

langue.

2.6.1 La didactique du français et l'évaluation sommative

Dans un texte intitulé « L'évaluation du travail des élèves » - considéré comme une référence par de nombreux auteurs consultés - Simard, Dufays, Dolz et Garcia-Debanc (2010) abordent la question de l'évaluation sommative en français et proposent d'en édicter les conditions de son efficacité. Ce dernier terme n'est déjà pas anodin. Il renvoie à une forme d'utilité, de rendement. En effet, l'évaluation sommative est ici bel et bien mise en relation avec les processus d'enseignement et d'apprentissage, tout en étant réduite, via les grilles de critères qui la médiatisent, à une finalité de production. Nous constatons cette même perception de l'évaluation sommative chez certains autres chercheurs comme Gagnon, Monnier et Dolz (2014), lorsqu'ils affirment, s'agissant des critères d'évaluation :

Si la pertinence et l'utilité des grilles sont discutées par certains chercheurs (Daunay, 1994), il n'en demeure pas moins que bon nombre d'entre eux (Simard, Dufays, Dolz & Garcia-Debanc, 2010) s'accordent à dire qu'elles permettent de stabiliser les critères, les indicateurs précis à utiliser pour apprécier le degré d'adéquation à la tâche d'écriture proposée, ainsi que la pondération à accorder aux différents critères sélectionnés. Elles constituent ainsi une aide non seulement pour fixer le seuil de réussite, mais également pour communiquer le résultat à l'élève (p. 2).

Par ailleurs, l'évaluation sommative, toujours selon Simard et ses collègues, devrait en plus porter sur des savoirs déclaratifs, convoqués pour être « utilisés » en même temps que des savoir-faire, dans des tâches non fragmentées, complexes, porteuses de sens et, si possible, « dotées d'une dimension épistémologique et/ou de plaisir » (p. 19). Autant de principes généraux qui, selon nous, font l'impasse d'une conceptualisation de l'évaluation, tant du point de vue des paradigmes dans lesquels elle s'inscrit (De Ketele, 2012 ; Figari, 2012, 2013)¹⁸, que des systèmes desquels découlent les outils exploités. Constat que l'on peut encore réaliser à la lecture des travaux de Tobola Couchepin, Sanchez Abchi et Dolz (2016), lorsqu'ils présentent leurs travaux sur les pratiques d'évaluation sommative de la production écrite au primaire : la finesse de leur analyse didactique est à nos yeux péjorée par une absence de conceptualisation des grilles critériées utilisées comme outils évaluatifs¹⁹.

¹⁸ Par souci de clarté et pour éviter de mélanger les niveaux de problématisation, nous abordons l'influence des paradigmes sur les outils d'évaluation plus loin.

¹⁹ Ce constat est d'autant plus interpellant sachant que cette recherche fait partie du recueil de textes publié par l'AIRDF mentionné plus haut.

Enfin, notons que l'incomplétude conceptuelle de l'évaluation sommative présente dans ces approches didactiques est d'autant plus intrigante qu'elle fait par ailleurs systématiquement l'impasse sur la construction de la note. Dans les sources consultées, les auteurs proposent des modèles de grilles de critères qui tiennent compte d'une forme de pondération - par ailleurs trop peu problématisée encore une fois - mais le pas jusqu'à l'acte de noter n'est pas développé. C'est notamment une des raisons qui nous fera revenir sur cette question plus loin en abordant justement les recherches qui ont documenté ce processus complexe qui est fortement dépendant, selon nous, de l'acte d'évaluer.

Un constat identique peut encore être réalisé à la lecture de travaux qui ont abordé l'évaluation d'objets de savoir spécifiques. Par exemple, en 2001, Lafontaine fait un tour d'horizon des travaux ayant abordé la question de l'évaluation de la lecture. On n'y trouve aucune trace d'une quelconque thématisation de l'évaluation sommative ni de la notation. En 2014, Roth, de Pietro et Sanchez Abchi abordent également la question de l'évaluation de la lecture, mais plus spécifiquement en référence au PER. Bien que mettant l'accent sur le rapport étroit devant exister entre enseignement et évaluation, nous observons les mêmes incomplétudes²⁰. En 2016, la revue de didactique et de pédagogie du français *Recherches* porte sur l'évaluation²¹. Les éditorialistes y revendiquent une approche critique que certains auteurs thématisent au niveau de l'évaluation sommative. Par exemple, Mottier Lopez, Serry et Sales Cordeiro analysent la production de l'écrit du fait divers au primaire. Si leurs propos sont très pertinents jusqu'à la pondération, rien n'est dit sur la construction de la note. Enfin, la même année, un ouvrage important sur le jugement professionnel²² rend compte de recherches dont certaines ont pour ancrage l'évaluation sommative. On peut y lire notamment le texte de Goasdoué, Vantourout et Bedoin (2016) qui s'intéresse aux composantes du jugement professionnel lors de la correction de dissertations en français, mais qui ne va pas non plus jusqu'à investiguer la dimension pragmatique de la construction de la note. Par exemple, lorsque les auteurs disent que « la note n'est pas la somme des jugements produits lors de la correction. Des appréciations semblables peuvent conduire à des notes divergentes et inversement des appréciations divergentes peuvent se solder par des notes voisines » (p. 75), on souhaiterait en savoir

²⁰ Notons par ailleurs que les auteurs utilisent indifféremment « évaluer » et « mesurer » sans les définir préalablement.

²¹ C'est la quatrième fois que cette revue aborde cette problématique, après les numéros 6 (1987), 21 (1994) et 38 (2003).

²² « Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages », coordonné par Lucie Mottier Lopez et Walther Tessaro, et publié chez Peter Lang.

davantage sur les outils auxquels ils se réfèrent, ou sur une éventuelle conceptualisation des pratiques de pondération et/ou de notation.

2.6.2 La didactique des mathématiques et l'évaluation sommative

Du côté de la didactique des mathématiques, force est de constater que nos recherches dans les travaux francophones sur les pratiques évaluatives sommatives des enseignants ont été également peu fructueuses. Pourtant, deux textes sont jugés fondateurs par la communauté scientifique. Le premier est de Chevallard (1989), dans lequel il montre le lien qui réunit l'acte d'évaluer et l'objet évalué :

L'évaluation, le « test » par lequel elle se réalise, plutôt qu'une mesure fournit *un écran sur lequel vient se projeter l'objet évalué* – le rapport de l'élève à l'objet de savoir, ou de tout autre catégorie d'acteurs qu'on voudra à tout autre objet qu'on voudra. Une telle image est potentiellement véridictrice : elle nous dit, ou du moins *on pourra chercher à y lire*, une certaine vérité sur ce rapport, sur cet objet que l'on évalue. (p. 16).

Il est ici intéressant de constater que l'auteur distingue déjà ici, certes dans des cadres anthropologiques, l'acte de mesurer et implicitement celui d'évaluer. Nous y reviendrons en détail plus loin. La pertinence de mettre l'objet au centre du processus évaluatif est aussi bel et bien là, mise en perspective avec cette question de vérité que nous assimilons aux enjeux de validité et/ou de cohérence, omniprésents. On observe donc, dans cette référence incontournable, une volonté didactique de problématiser l'objet d'évaluation. Toutefois, ici aussi, la question du verdict sur l'objet, via une note, n'est pas abordée par l'auteur. Nous y reviendrons également.

Le second apport est de Martinand (1986). L'auteur pose les bases de ce que pourrait être un regard didactique sur l'évaluation. Il met en évidence les potentiels risques de considérer l'évaluation à l'aune unique des objectifs²³. Pour lui, il est fondamental de mettre au centre du processus évaluatif une réflexion conjointe sur les objectifs et sur les objets. Dans ce sens, il accorde au contenu d'enseignement « une importance que les généralistes des sciences de l'éducation avaient un peu trop tendance à oublier ou même à nier » (Geminard & Martinand, 1987, p.114). Il est alors surprenant de constater que du point de vue de l'étude des pratiques évaluatives sommatives, la didactique des mathématiques est depuis restée discrète, à notre

²³ Nous faisons l'hypothèse ici que les propos de l'auteur peuvent être transposés à la notion de compétences, pas encore en vogue dans le monde scolaire au moment où il publie son texte.

connaissance. Et, paradoxalement, ce sont des recherches non didactiques qui fournissent des résultats intéressants.

Centrées sur la discipline, certaines ont montré, par exemple, que les enseignants de mathématiques se trouvent face à une complexité très élevée lorsqu'il s'agit de critérier des épreuves (Meier et al., 2006), ou ont documenté les enjeux à produire des jugements globaux dans cette discipline (Klein et al., 2009). De manière complémentaire, d'autres résultats de recherche sur les pratiques apportent des connaissances intéressantes. Par exemple, en France, Braxmeyer et al. (2005) révèlent que les évaluations dans cette discipline au secondaire sont le plus souvent longues et écrites, proposent des tâches de niveau de difficultés et de nature variés. Ce sont régulièrement des tâches de restitution à l'image de celles des manuels. Mais de manière générale, ils constatent que « les enseignants de mathématiques interrogés ont eu un discours assez sommaire autour de l'évaluation » (p. 29). Toutefois, ils précisent que l'« on peut percevoir des signes indubitables de pratiques évaluatives relevant aussi d'autres perspectives que la seule mesure des acquis des élèves et de l'appréciation de leurs progrès : en effet, comprendre la nature des erreurs concerne 79,4 % des professeurs de mathématiques » (p. 68). Notons enfin qu'une recherche par mots clés d'articles portant sur l'évaluation sommative et sur la notation réalisée sur le site de la revue *Recherches en didactique des mathématiques* (RDM) n'a donné aucun résultat.

À la lumière des travaux recensés dans cette discipline, nous observons qu'il semble exister un manque dans la recherche francophone²⁴ sur les pratiques évaluatives sommatives réelles des enseignants articulant une réflexion sur les objets disciplinaires et une conceptualisation de l'évaluation, jusqu'à la notation. Les derniers travaux de Sayac (2018), en France, tentent de combler le vide. Toutefois, s'ils éclairent de manière pertinente la problématique de la complexité des tâches, ils la séparent en revanche des pratiques de pondération et de notation, alors que selon nous, elles sont indissociables.

Cette situation contraste avec la recherche en évaluation sommative au sein de la communauté anglophone, qui a une longue tradition en la matière (McGatha & Bush, 2013). En effet, on voit encore une fois nombre de chercheurs inscrire leurs travaux dans cette discipline (e. g., Black et al. 2010 ; McKinney, Chappell, Berry & Hickmann, 2009 ; Smith, Smith & De Lisi, 2001 ;

²⁴ La recherche dans le monde anglo-saxon est en revanche très prolifique à ce niveau (voir par exemple le site de la revue *Educational Studies in Mathematics* : <https://link.springer.com/journal/10649>).

Stiggins, 2001). Cette recherche, durant les vingt dernières années, s'est centrée sur deux axes : l'étude des pratiques réelles et le jugement des enseignants. Le premier axe a permis de montrer que les pratiques évaluatives sommatives sont dans cette discipline souvent traditionnelles (Ohlsen, 2007), le second a mis en évidence l'importance des connaissances de l'objet de savoir dans la pertinence des jugements (Martinez et al., 2009).

Cela étant posé, il s'agit maintenant de passer en revue ce qu'offre la littérature en termes de résultats relativement à la notation, car notre problématique de recherche considère - à l'instar de certains chercheurs comme McMillan et Nash (2000) ou Walwoord et Johnson Anderson (2009) pour ne citer que ceux-ci - que les pratiques d'évaluation sommative et de notation sont en forte interdépendance.

3 La note : ou la nécessité de reformuler le débat qu'elle suscite

Une abondante littérature montre l'impact négatif que la note a sur les apprentissages et la motivation des élèves, dénonce son utilisation au service du classement des individus, la condamne puisqu'elle est un outil au service d'une logique méritocratique (e.g., Butera, Buchs & Darnon, 2010 ; Maulini, 2012 ; Merle 1998, 2007). Certains auteurs issus de l'Éducation nouvelle, comme Neumeyer et Vellas (2015), ont « la conviction qu'il nous faut au plus vite nous "désintoxiquer" de la note car elle contribue à la ségrégation scolaire, alimente le tri et l'exclusion sociale » (p. 8). Leur collègue Pepinster (2015) parle d'addiction à la note : les enseignants seraient « des malades de la note, à leur insu le plus souvent » (p. 61). D'autres chercheurs réduisent la note à une « transaction commerciale » (Barlow (2003, p. 45), à une « marchandisation » (De Vecchi, 2011, p. 71), avec l'argument notamment qu'elle n'est pas en mesure de dire quoi que ce soit d'intéressant de et à l'élève.

Relativement à notre problématique, nous faisons l'hypothèse que cette littérature investit un débat d'arrière-garde, à nos yeux, obsolète. Un débat qui parfois même engage les acteurs dans des échanges peu féconds²⁵, comme en France, où cette question est devenue un axe central de la réforme de l'école initiée en 2015.

Certes, il n'est ici pas question de remettre en question les multiples études sur la pertinence des notes menées par les docimologues (e. g., Noizet, 1961; Piéron, 1963; Noizet & Caverni,

²⁵ Voir notamment : http://www.lemonde.fr/societe/article/2015/09/29/reforme-de-l-evaluation-ce-que-prepares-najat-vallaud-belkacem_4776089_3224.html.

1978), qui ont mis en évidence les biais qui affectent les notes et leur impact négatif sur les parcours scolaires des apprenants. Toutefois, pour préciser notre positionnement, nous nous référerons à des travaux qui se distancient de l'étude de pratiques de notation normatives, comme ceux de Butera, Buchs et Darnon (2010). Les références que nous détaillons ci-après, plus constructives selon nous, visent à mettre en débat quatre axes de réflexion qui nous semblent aujourd'hui centraux dans tout travail de recherche dans ce champ, et à ce jour insuffisamment développés dans les contributions recensées : a. il y a une urgence à (re)définir la finalité de la notation, b. le problème est moins la note que sa construction, c. la note peut relever d'un raisonnement pédagogique, et d. à certaines conditions, la note est en mesure de fournir des renseignements sur les apprentissages des élèves. Reprenons-les dans l'ordre.

3.1 De l'urgence à (re)définir la finalité de l'évaluation sommative notée

Dans un ouvrage pour nous marquant, Hadji (2012) pose un regard sociologique, voire philosophique, sur la place de l'évaluation dans notre contexte actuel, en explicitant les dérives de cette dernière lorsqu'elle aboutit à une note. Son constat est sans concession : l'évaluation est devenue une obsession. Il montre, au travers de nombreux exemples, qu'elle est omniprésente, prend des formes multiples, est idéologique car elle véhicule des valeurs. Située socialement, mais aussi culturellement, politiquement et historiquement, il postule que les acteurs, de manière générale, entretiennent avec elle un rapport ambigu : ils la détestent, et en même temps, ils y adhèrent. La raison selon lui réside dans le fait que l'évaluation - et de surcroît lorsqu'elle est notée - fait une double promesse. L'une est rassurante : elle donne des repères, un cadre temporel, elle fixe des objectifs. L'autre est narcissique : l'évaluation permet de s'améliorer, d'être le meilleur, potentiellement, un jour. Cette tension relationnelle à l'évaluation serait à la base de notre passivité face à ce raz de marée évaluatif. Plus loin, Hadji explicite trois problèmes majeurs en évaluation :

- On évalue trop (on ne réfléchit plus à l'objet que l'on souhaite évaluer) ;
- On évalue trop souvent mal (il manque des connaissances chez les évaluateurs notamment au niveau des méthodologies d'évaluation) ;
- On évalue à des fins contestables (on ne sait pas pourquoi l'on évalue, quelle est la finalité de la démarche, ou pour des raisons peu honorables).

L'évaluation devient alors un outil au service d'une entreprise qui la dépasse, et pour laquelle elle va servir de leurre, d'alibi ou de prétexte. Ces éléments nous paraissent particulièrement éclairants pour comprendre comment nous abordons la problématique de la notation.

3.2 Le problème : la construction de la note

Selon nous, il s'agit ici prioritairement, comme le précise toujours Hadji (2012), d'analyser « la “qualité” des conditions de production de la note » (p. 172), et non la pertinence de son existence, au-delà de la fascination du chiffre qu'elle véhicule et de la dictature des moyennes qu'elle convoque la plupart du temps. Il n'est dès lors plus question de questionner sa véracité ni sa légitimité, mais d'analyser les pratiques de construction des notes des enseignants, afin, notamment, de comprendre comment ils s'y prennent et de voir quelles finalités ils poursuivent en notant les apprentissages des élèves.

Cette nécessité de s'attarder sur les processus de notation, relevant d'une pratique très complexe, est soulevée par de multiples auteurs. Nous savons en effet aujourd'hui que les pratiques de notation varient souvent en fonction d'un certain nombre de facteurs comme la matière (Resh, 2009), le niveau des élèves (Guskey, 2009) ou leur milieu socio-économique (Klapp Lekhom, 2011), et que bien entendu elles sont intimement liées à l'évaluateur (Kitabgi, 2009) ou à la culture de l'établissement (Duru-Bellat, 2010). Il s'agit donc de tenter une objectivation de la notation. Dubus (2006), par exemple, parle d'une notation raisonnée. Il propose une démarche compréhensive visant à « s'intéresser aux questions de l'élaboration des notes, et des différentes techniques de manipulation, agrégation, conservation qu'on leur fait subir, ainsi que des genres de décisions auxquelles ces techniques conduisent » (p. 11). On perçoit ici une perspective dynamique qui est aussi la nôtre, et que relayent d'autres travaux. Par exemple, McMillan et Nash (2000) définissent le processus de notation comme une étape clé de l'acte évaluatif sommatif. Walvoord et Johnson Anderson (2009) montrent le rapport qui existe entre enseignement, apprentissage, évaluation *et* notation, ce qui signifie que l'on ne peut noter sans se référer à l'apprentissage qui a été dispensé et évalué. Cross et Frary (1999) affirment qu'il y a urgence à distinguer les pratiques de notation défendables des autres, et dans une perspective assez pragmatique, s'entendent à formuler l'existence de bonnes pratiques qu'il s'agit urgemment de développer en formation²⁶. Autant d'éléments qui nous encouragent, dans notre cadre conceptuel, à donner à ce champ de recherche l'importance qu'il mérite. Et ceci même si la note a encore bien mauvaise presse dans certaines communautés de chercheurs, qui la considèrent comme un objet d'étude peu intéressant, car faussement vue comme basement

²⁶ Afin de ne pas alourdir cette partie de l'exposé, nous renvoyons ce raisonnement sur les « bonnes pratiques » à une conception d'une notation alignée d'un point de vue curriculaire que nous développons plus loin, puis théoriquement dans notre cadre conceptuel (Partie 3).

technique (Mottier Lopez, 2015, correspondance personnelle).

3.3 Une note pour jouer trop de rôles

En rapport avec le point précédent et fidèlement à notre problématique, il nous importe de questionner la référence autour de laquelle la note est élaborée. Si elle est construite en référence à une norme – le plus souvent la moyenne des points obtenus par les élèves – elle émerge d'une distribution normale par rapport à cette moyenne, et représente donc le résultat d'une comparaison entre élèves. Cardinet (1990) situe l'émergence de cette démarche dans le courant de la psychométrie, alors obsédée par le besoin de tout quantifier, et dont la devise est : « tout ce qui existe dans une certaine quantité, que l'on peut donc mesurer » (p. 1). Cette perspective appliquée à l'évaluation des apprentissages, unanimement critiquée par les auteurs consultés (e.g., Antib, 2003 ; Brookhart, 2005 ; Dubus, 2006 ; Mottier Lopez, 2015) a fait - et fait encore - des ravages sur les parcours de quantités d'élèves. Par exemple, Crahay (2010) montre qu'un élève ayant beaucoup progressé dans un groupe classe « fort » est jugé plus sévèrement que si son groupe d'appartenance avait été « moyen ».

Par ailleurs, des travaux montrent que la notation revêt des enjeux qui dépassent souvent ce pour quoi elle est conçue. L'un d'entre eux, comme l'expliquent Bressoux (2002) ou Merle (2007), est le maintien d'une interaction vivable en classe entre l'enseignant et les élèves où le premier peut garder son rôle de détenteur du savoir, et les seconds percevoir qu'ils peuvent apprendre avec lui. D'autres auteurs voient encore dans la note le levier d'une forme de pouvoir qui assoit l'autorité de l'enseignant, récompense ou sanctionne (Desclaux, 2014). Elle est également, pour certains professionnels, au service d'une motivation extrinsèque (Black et al., 2010). Elle conforte également bon nombre de croyances. Par exemple, Merle (2007) montre que :

« L'adhésion professorale à l'idée de stabilité des compétences scolaires d'une année à l'autre existe : le mauvais élève est voué à le rester, le brillant à confirmer sa réussite, celui dont le grand frère était « médiocre » va probablement marcher dans ses pas, etc. ». (p. 12)

De son côté, Kitabgi (2009) affirme que la note semble surtout « servir à s'assurer du maintien d'un niveau de travail scolaire élevé » (p. 37). Tous ces phénomènes sont bien évidemment le fruit de mécanismes le plus souvent inconscients. Ils se conjuguent fréquemment, donnant aux processus de construction de la note une forme de stabilité dans la pratique. Ceci expliquerait pourquoi il est très difficile de la remettre en question. En effet, grâce à des travaux dans le

champ, nous connaissons le rapport ambigu que les enseignants entretiennent avec la note (Brookhart, 2013 ; Dubus, 2006 ; Kitabgi, 2009 ; Winger, 2009), puisque d'un côté elle est leur seul moyen de « rendre compte » de leur travail et de l'autre, elle est un outil, au quotidien, qui permet de stabiliser les conditions de l'apprendre ensemble.

Malgré ces biais, une littérature montre que la note, si elle est construite au travers d'un raisonnement pédagogique, peut être en mesure de fournir des informations pertinentes sur les apprentissages des élèves.

3.4 Une urgence supplémentaire : re-considérer la note en référence aux apprentissages

La note fait donc partie de l'habitus de l'enseignant (Gimonnet, 2007). Envisager qu'elle puisse être construite en référence aux apprentissages participerait notamment à abolir la perspective normative en évaluation sommative, à proposer d'autres outils, à remettre au centre du raisonnement évaluatif les objets de savoir et une acuité portée sur les progrès des élèves, au travers de démarches de formation visant la modification des postures et des pratiques d'évaluation sommative. Sacré défi ! En effet, comme le précisent Mottier Lopez et Laveault (2008), « le fait même de situer l'élève par rapport à ses apprentissages plutôt que par rapport à ses pairs constitue une révolution copernicienne caractérisée par une évaluation critériée plutôt que normative » (p. 8). Cette littérature va donc dans le sens d'une de nos questions : à quelles conditions cette révolution des pratiques de notation est-elle possible ?

On le voit, les éléments cités jusque-là sont fondamentaux dans la tentative de mieux décrire et comprendre les pratiques de notation, situées idéologiquement et historiquement. En suivant Mottier Lopez et Figari (2012), nous faisons l'hypothèse qu'ils appellent à penser une cohérence de l'évaluation sommative à l'aune de modèles théoriques, afin de sortir de pratiques souvent trop intuitives. Nous avons choisi dans notre recherche le modèle de l'alignement curriculaire : son mode de structuration, son rapport étroit avec les pratiques réelles ainsi que l'abondante littérature dont il est l'objet nous ont encouragé à le considérer comme un modèle pouvant être au service de la compréhension de la pratique évaluative sommative et de la notation.

4 Un modèle²⁷ pour appréhender la cohérence de l'évaluation sommative : l'alignement curriculaire

La question de la cohérence de la pratique évaluative a mobilisé bon nombre de chercheurs (e. g., Bateman et al., 2009 ; Biggs, 2003 ; Harlen, 2005, 2012 ; Royce Sadler, 2009) qui ont chacun à leur manière montré le rapport étroit qui existe entre les processus d'enseignement, d'apprentissage et d'évaluation. Mottier Lopez et Laveault (2008) expriment cet intérêt de la manière suivante : « Cette recherche de cohérence et d'articulation des trois domaines est tout à fait caractéristique des évolutions les plus récentes de l'évaluation en éducation » (p. 9). On voit donc ici que cette question offre des perspectives de recherche et de formation larges.

La littérature qui traite spécifiquement du modèle d'alignement curriculaire repose sur le même postulat. Elle est issue d'un champ de la recherche bien circonscrit : celui des curricula. Ce modèle soutient donc l'idée que toute pratique d'évaluation est en forte cohérence avec les objectifs prescrits, l'enseignement dispensé, les tâches d'apprentissage proposées aux élèves et les activités effectivement réalisées par ces derniers (e. g., Anderson, 2002 ; Biggs, 1999, 2003 ; Gauthier, Mellouki, Bissonnette & Richard, 2005 ; Martone & Sireci, 2009). L'enjeu pour l'évaluation est ici de porter sur les apprentissages prévus et effectifs. Dans ce sens, le modèle implique de considérer l'élaboration d'épreuves évaluatives sommatives comme un processus complexe, dynamique, spécifique et centré sur les apprentissages. À ce titre, la littérature consultée dans ce champ est univoque à deux niveaux. Tout d'abord, la notion d'*alignement* renvoie à une conceptualisation de la cohérence entre les différentes dimensions du modèle élaborée à l'aune d'une taxonomie d'habiletés cognitives appariées à des contenus. Ensuite, la dimension *curriculaire* renvoie au fait que cette cohérence se doit d'être ancrée dans un curriculum.

4.1 L'alignement curriculaire : un réseau conceptuel

En 2005, Gauthier et ses collègues publient un état de la recherche sur les écoles efficaces et la réussite scolaire des élèves à risque en Amérique du Nord. En s'appuyant sur l'analyse de nombreux travaux, ils montrent que « l'alignement curriculaire apparaît grandement susceptible d'améliorer la qualité de l'enseignement et l'efficacité des écoles » (p. 28)²⁸. Certes,

²⁷ Nous insistons sur la distinction entre modèle et modélisation dans notre cadre conceptuel (Partie 3).

²⁸ À l'aide du moteur de recherche ERIC, ils relèvent que de 1966 à 2005, 245 recherches ont été publiées sur ce thème, mais qu'aucune méta-analyse n'a toutefois été réalisée.

sa mise en œuvre ne garantit pas à elle seule et *de facto* de meilleurs apprentissages. Mais il demeure que c'est un des facteurs de la réussite, l'une des clés de voûte du succès de toute réforme en éducation (Brophy, 2000).

Dans la littérature, l'alignement curriculaire est problématisé au regard de son *mode de structuration*. Deux niveaux principaux peuvent être différenciés. Le premier investigate la cohérence qui existe entre les pratiques d'enseignement, les évaluations internes à la classe et les « standards » externes (e. g., Porter, 2002 ; Squires, 2009). Les travaux produits montrent notamment le peu de validité qu'ont les évaluations externes si elles ne sont pas en cohérence avec les processus d'enseignement et d'apprentissage et les pratiques évaluatives réelles des enseignants.

Le second *mode de structuration* de l'alignement curriculaire analyse à quelles conditions l'évaluation interne, d'un point de vue général, peut être en cohérence avec les processus d'enseignement et d'apprentissage et le curriculum. C'est ce second *mode de structuration* qui nous intéresse, notre problématique de recherche étant centrée sur les pratiques effectives d'évaluation sommative d'enseignants en exercice. Dès lors, puisque la littérature relative au modèle met en lien les objectifs prescrits et les tâches évaluatives avec les démarches d'évaluation, nous nous orienterons vers une série de recherches qui traitent de curriculum, de référentiel, d'objectifs, d'objets, de tâches. Une attention particulière sera accordée à la notion de cohérence d'alignement curriculaire, centrale, qui mérite selon nous une attention particulière. Puis, nous aborderons la littérature sur la pondération et la notation pour concrétiser ce qui se joue lorsque l'on parle d'évaluation sommative au sein du modèle, en documentant, au travers de recherches significatives, à quelles conditions ces éléments entretiennent des liens de cohérence en référence au modèle de départ. Ce travail nous amènera à concevoir une acception « élargie » du modèle.

4.2 Un modèle linéaire

Un des principaux chercheurs à avoir développé le modèle de l'alignement curriculaire au cœur des pratiques enseignantes est Biggs²⁹ (1999, 2003, 2010). Conjuguant une approche des apprentissages qu'il nomme cognitiviste avec le principe que ces derniers doivent se référer à des objectifs prescrits, il modélise le rapport de cohérence qui existe entre les processus

²⁹ Ses travaux portent sur l'enseignement supérieur, mais leurs apports sont transposables à la scolarité obligatoire, comme l'ont montré une partie des recherches citées dans cette partie.

d'enseignement et d'apprentissage, idéalement référencés à un curriculum, et les pratiques évaluatives sommatives. Selon lui, « the teaching methods used and the assessment tasks, are *aligned* to the learning activities assumed in the intended outcomes » (2003, p. 13). Dès lors, l'alignement curriculaire est un système dont les composantes (objectifs du curriculum, tâches d'apprentissage et démarches d'évaluation sommative) interagissent étroitement.

Dans une logique pragmatique, il définit les quatre étapes qui permettent cet alignement :

- La définition des objectifs visés ;
- Le choix et la conception des tâches d'apprentissage pour les travailler au cœur des processus d'enseignement ;
- La phase d'évaluation, à prévoir en relation avec les deux points précédents ;
- La production de la note, elle-même en congruence avec les apprentissages visés et réalisés.

Les éléments principaux et essentiels de l'alignement curriculaire que nous allons considérer dans le cadre de notre recherche, bien qu'à certains égards insuffisamment théorisés par l'auteur selon nous, sont présents. Nous retrouvons également ici deux étapes clés de toute démarche évaluative sommative : les pratiques d'évaluation et les pratiques de notation. Nous observons sur ce point que Biggs, comme par exemple Anderson (2002), les met en étroite interdépendance, tout en les distinguant et en passant rapidement sur leurs caractéristiques propres. Nous y reviendrons plus loin, et très spécifiquement dans notre cadre conceptuel, puis dans nos chapitres de résultats. Enfin, nous observons que cette linéarité n'est que peu remise en question et qu'elle s'appuie sur la notion de cohérence.

4.3 Une taxonomie pour lire la cohérence d'alignement curriculaire

Dans la littérature sur l'alignement curriculaire, les pratiques évaluatives et les objectifs du curriculum entretiennent des rapports de cohérence conçus d'un point de vue d'habiletés cognitives et de contenu. Par exemple, enseigner l'analyse de texte demande des pratiques quotidiennes en classe sur cet objectif et cet objet de savoir, et il doit en aller de même pour les démarches d'évaluation. C'est la raison pour laquelle les chercheurs consultés parlent d'une cohérence d'alignement curriculaire construite à l'aide d'une taxonomie (e. g., Anderson, 2002 ; Biggs, 2003).

Plusieurs formes de taxonomies existent dans les études scientifiques, exploitées comme des

outils compréhensifs. Mais c'est au niveau des taxonomies cognitives que la recherche a été la plus productive, avec diverses tentatives de hiérarchiser la complexité des habiletés cognitives en jeu dans tout processus d'apprentissage (e. g., Biggs, 2003 ; Marzano, 2001). Nous n'avons toutefois pas relevé de consensus sur une taxonomie dans les travaux consultés. En effet, selon leur ancrage épistémologique et leurs méthodologies, les chercheurs opèrent des choix différents pour caractériser spécifiquement l'apprentissage, ce qui prouve que la détermination d'une taxonomie traduit un rapport particulier aux conceptions que l'on a de ce dernier (Pellegrino, 2006), voire à la discipline concernée³⁰. La littérature que nous retenons ici est celle qui a problématisé cette cohérence curriculaire à l'aide de la taxonomie d'Anderson et Krathwohl (2001)³¹, et ceci, pour trois raisons.

La première raison qui justifie notre choix de retenir cette taxonomie vient du fait que ces chercheurs l'ont construite sur la base de celle de Bloom (1969, 1971), qui a longtemps fait référence, mais qu'ils ont actualisée au regard de l'avancée des recherches sur la cognition. Bloom, dans une perspective alors behavioriste, avait théorisé l'activité mentale en six paliers, successifs et montant en complexité, afin de fournir un outil théorique pour appréhender au mieux l'apprentissage (connaître, comprendre, analyser, synthétiser et évaluer). Cet outil, visant à étayer les pratiques réelles, a depuis montré ses limites, puisque nous savons aujourd'hui que l'on n'apprend pas par paliers ou par couches (on peut être capable de comprendre un phénomène tout en ne connaissant pas par cœur toutes ses dimensions), mais de manière systémique.

La deuxième raison est que d'autres chercheurs se sont appuyés sur cette taxonomie pour produire des résultats intéressants en partant de l'étude des pratiques d'enseignement-apprentissage et d'évaluation sommative. Par exemple, au niveau du CEGEP au Québec, Bateman et ses collègues (2009) ont mis en évidence, à objectifs communs, les forts écarts qui existent dans les pratiques d'enseignement en salle de classe, et surtout, dans les pratiques évaluatives sommatives.

Enfin, la perspective cognitiviste de cette taxonomie nous semble pertinente pour caractériser la cohérence curriculaire dans les pratiques évaluatives sommatives. Dans cette perspective, l'apprentissage y est conçu comme la capacité à maîtriser des objectifs formulés en termes

³⁰ Nous pensons notamment à la taxonomie élaborée par Bodin (2010) pour les mathématiques, inspirée également de celle d'Anderson et Krathwohl (2001).

³¹ La taxonomie est disponible au point 1.2.8 de la partie 6 des annexes.

d'habiletés cognitives appariées à des contenus. Les auteurs qui suivent cette perspective accordent donc de l'importance, du point de vue de la cohérence, à deux éléments clés : la complexité de l'habileté cognitive en jeu, et conjointement, les caractéristiques de l'objet disciplinaire (Hammerness, 2006). L'objectif d'apprentissage n'est donc pas ici décrit par son niveau de complexité uniquement. Il n'aboutit pas non plus à un raisonnement linéaire qui l'enfermerait dans une méthode d'enseignement et d'apprentissage donnée. Il peut dès lors être défini comme une capacité comprenant une dimension générale et spécifique. La première dimension renvoie à l'activité mentale de l'élève (se souvenir, analyser, différencier, etc.), ou aux habiletés cognitives de la taxonomie. La seconde se rapporte au contenu sur lequel porte l'apprentissage en question : la relative en français, les isométries en mathématiques, etc. Ce contenu fait référence aux caractéristiques des objets spécifiquement disciplinaires (Marton & Tsui, 2004). Il est d'ailleurs intéressant ici de constater que Cardinet, en 1984 déjà, mettait l'objet de savoir au centre des réflexions. Il posait clairement, en précurseur, cette « cohérence nécessaire entre objectifs, enseignement, évaluation et remédiation » (p. 42).

Cette problématique de la cohérence n'est donc pas nouvelle. La centration toutefois sur l'idée d'une cohérence *curriculaire*, en termes d'alignement, qui insiste spécialement sur les relations qui doivent exister entre toutes les composantes du modèle, du curriculum à la notation, est pour nous novatrice. En effet, elle responsabilise l'enseignant qui devient expert des contenus à enseigner (Klenowski & Wyatt Smith, 2014), afin qu'il puisse avoir conscience de la complexité des apprentissages visés pour les décrire. En d'autres termes, cela signifie que le choix de la taxonomie reflète la complexité des contenus qui vont être évalués (Gagné, Dumont, Brunet & Boucher, 2013). Sur ce point, les auteurs consultés sont d'accord : cette démarche est d'une grande ampleur, et un enseignant tout seul aurait du mal à la réaliser (Li, Klahr & Siler, 2006). D'où la nécessité d'un travail de concertation pour déterminer les critères d'alignement curriculaire d'un point de vue d'évaluation sommative en référence à une taxonomie (Webb, 2007).

4.4 Une cohérence « *top down* » et « *bottom-up* »

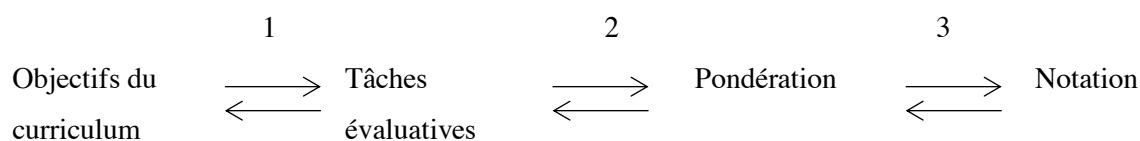
La cohérence de l'alignement curriculaire, telle que partagée dans la littérature, peut se penser à partir de chaque composante du modèle (enseignement, apprentissage ou évaluation). La plupart des auteurs ayant travaillé sur la question (e. g., Walvoord & Johnson Anderson, 2009) la thématisent dans sa perspective « *top-down* » (p. 18), c'est-à-dire du curriculum jusqu'aux tâches d'évaluation. Là, il s'agit pour l'enseignant d'affiner les objectifs, de les décliner en

unités d'apprentissages significatives afin de les concrétiser dans des tâches. Mais ce travail de mise en lien peut également se réaliser dans un mouvement ascendant, dans une perspective « *bottom-up* ». Par exemple, l'enseignant analyse des tâches évaluatives au regard de leur cohérence avec les objectifs du curriculum. Dans les deux perspectives, l'idée est de mettre au centre de la réflexion une inscription de la cohérence dans des contenus et des apprentissages référencés, jusqu'à la pondération et la notation. Mais que signifie une perspective « *bottom-up* » ou « *top-down* » considérée depuis la formulation de critères, ou à partir de l'élaboration d'échelles de notation ? Et dans quelle mesure ces perspectives sont-elles toujours pertinentes pour penser cette question de cohérence des pratiques évaluatives ?

5 Une acception de l'alignement curriculaire nécessairement élargie

Cette acception élargie renvoie à un enjeu théorique : mettre en évidence les différentes dimensions « cachées » de l'évaluation sommative, à savoir la pondération et la notation. En effet, dans toutes les sources consultées, rares sont celles qui problématisent ces deux phases du processus évaluatif. Pourtant, elles sont cruciales à nos yeux, et présentes dans l'immense majorité des contextes scolaires.

Au sein même des travaux sur l'alignement curriculaire également, peu nombreuses sont les recherches qui ont théorisé ces questions de pondération et de notation. Si certains auteurs (e. g., Royce Sadler, 2009) sont allés jusqu'à modéliser des manières très précises de formuler des critères et de construire des échelles alignées d'un point de vue curriculaire, nous avons été surpris de constater que leur conceptualisation au sein même du modèle est souvent peu approfondie selon nous, et surtout, peu contextualisée. Par exemple, peut-on penser la cohérence curriculaire de la même manière dans un système où il existe potentiellement onze notes³² et dans un autre où il n'y en a que quatre, dont une seule signifiant l'échec, comme en Finlande ou en Suède ? Dès lors, il nous semble fondamental d'étendre le modèle de l'alignement curriculaire au travers d'un développement conceptuel que sont ses deux dernières composantes. Nous le schématiserons ainsi :



³² Comme c'est le cas dans le canton de Vaud.

Figure 1 : Modélisation de l'alignement curriculaire élargi

Nous observons alors ici le rapport de cohérence réciproque que chaque composante entretient avec ses voisines, dans une perspective dialectique, d'où les flèches doubles. Cela étant, précisons l'importance d'inscrire cette modélisation dans les processus d'enseignement et d'apprentissage. Comme nous l'avons dit déjà, rendre compte d'un processus d'alignement en en faisant fi serait incohérent. De ce fait, et en nous référant aux pratiques réelles, nous devons considérer que parfois les composantes n'entretiennent pas, directement, un rapport « en cascade »³³. Par exemple, de nombreuses sources sur les critères (e. g., Andrade, 2005 ; Meier, Rich & Cady, 2006 ; Reddy & Andrade, 2010 ; Royce Sadler, 2005) montrent que la pondération est un acte intimement lié à une forme de caractérisation des objectifs, sans forcément passer directement par l'élaboration des tâches évaluatives. En effet, il n'est pas rare de voir des enseignants, tous cycles confondus, formuler des critères d'évaluation en se référant essentiellement aux objectifs qu'ils disent évaluer, sans prendre forcément en considération les consignes de travail.

Dans le cadre conceptuel, nous présenterons donc une nouvelle acception de l'alignement curriculaire cette fois-ci *élargie*, dans une perspective *systémique*, étant donné que c'est à l'aune des apprentissages menés en classe au quotidien que l'évaluation sommative devrait idéalement être organisée. Dans ce sens, nous questionnerons la conception étapiste que Biggs et ses collègues ont de l'*alignement* curriculaire, qui renvoie davantage à des actions structurées en termes de phases qu'à des processus de mise en lien permanents des différentes composantes du modèle. Pour cela, nous commencerons par documenter le rôle du curriculum dans le modèle de l'alignement curriculaire.

5.1 Les conditions pour un curriculum à des fins d'alignement curriculaire

La littérature consultée pose un certain nombre de conditions pour qu'un curriculum soit utilisable à des fins d'alignement. Constat intéressant : elles font toutes référence à la nécessité d'explicitier et de rendre visibles les contenus et les niveaux d'exigence des apprentissages, d'un point de vue de complexité et de contenus (e. g., Anderson, 2002 ; Biggs, 2003 ; Biggs & Tang, 2007, 2011 ; Squires, 2009). Le curriculum devrait par ailleurs aider les enseignants à connaître les niveaux des élèves relativement à l'enseignement dispensé et à l'évaluation mise en œuvre.

³³ Ce qui aboutira, en fin de recherche, à une nouvelle perspective systémique du modèle présentée dans notre conclusion (Partie 9).

Ces éléments nous obligent donc à considérer qu'aucun curriculum, *de facto*, n'offre la possibilité de penser cet alignement curriculaire. Dans ce sens, la littérature montre qu'avant de penser à un alignement potentiel, il s'agit de définir ce qu'est, précisément, un curriculum, et en quoi il va permettre, ou non, un raisonnement aligné curriculairement.

Les sources consultées s'accordent à dire que l'enjeu majeur, dans une perspective d'alignement, est de construire une cohérence curriculaire entre le contenu du curriculum, l'enseignement et les tâches évaluatives (Hermann, Webb & Zuniga, 2007). Comme le précise Muller (2006), il s'agit pour l'enseignant de « se préoccuper, en premier lieu, de la dynamique des relations entre les “entités” composant un curriculum » (p. 107). C'est sans doute un premier pas. Dans notre contexte romand par exemple, il s'agit d'analyser dans quelle mesure le Plan d'Études Romand (PER) est un curriculum, et corollairement, à quelles conditions il peut jouer le rôle de référentiel d'évaluation pour les enseignants. L'enjeu est donc de taille. Ceci d'autant plus que Mottier Lopez, Tessaro, Dechamboux et Morales Villabona (2012) ont montré que l'usage du PER par les enseignants « est indirect pour définir les contenus et les objectifs de leurs évaluations certificatives internes, ce qui représente un décalage avec les recommandations institutionnelles qui incitent les enseignants à utiliser le plan d'études comme référence première » (p. 8).

Certains travaux, comme ceux de Carulla, Corti, de Pietro, Kassam, Roth, Sanchez Abchi et Singh (2013), ont tenté de problématiser cette tension. Ils mettent notamment en évidence l'importance de différencier référentiel et curriculum. Leurs caractéristiques sont-elles communes ? Quels éléments les différencient ? Et quel impact cela a-t-il sur les pratiques évaluatives des enseignants, en référence à ces repères ? De leur côté, Cros et Raisky (2010) présentent le référentiel comme « un outil de médiation normatif permettant aux activités humaines de s'y référer pour étudier un écart ou des différences » (p. 107). Son rôle serait alors essentiellement de situer l'individu par rapport à une norme, dans un processus actionnel dynamique. Mais de quelle norme parle-t-on dans un contexte scolaire, et qui plus est relativement à l'évaluation sommative ? Ceci d'autant plus si nous acceptons qu'un référentiel d'évaluation serait un « dispositif permettant de délivrer le diplôme avec ses types de contrôles et ses modalités » (Cros & Raisky, 2010, p. 109). Dans ce questionnement foisonnant, un élément fait toutefois consensus : il est admis qu'un curriculum ne peut être assimilé à un référentiel d'évaluation (Carulla & al., 2013 ; Figari & Remaud, 2014).

La différence entre curriculum et référentiel d'évaluation est donc fondamentale lorsque l'on

parle d'évaluation sommative. En effet, s'il est admis que « les concepteurs du PER ont choisi de présenter le référentiel d'enseignement sous forme de curriculum » (Marc & Wirthner, 2013, p. 21), rien, à notre connaissance, n'a été formalisé pour décrire ce plan d'études comme un référentiel d'évaluation. Au contraire, les débats sont en cours, puisque suite à leurs rapports, Carulla et al. (2013) précisent qu'il est temps d'aller voir du côté des pratiques évaluatives sommatives des enseignants pour penser ces questions de curriculum, mais surtout de référentiel. Nous souscrivons à ces propos, car diverses études ont montré que les pratiques évaluatives des enseignants en référence au curriculum sont loin de correspondre aux objectifs prescrits (e.g., Bateman et al., 2009 ; Grisay, 1984, 1988c, 1989). On peut faire l'hypothèse, selon les sources consultées, que ce flou serait notamment dû à des curricula peu clairs, voire éloignés des pratiques, ou, chez les enseignants, à des habitudes de penser l'évaluation en se référant essentiellement à ce qui a été fait en classe plutôt qu'à ce qui y a été appris. Dans ce sens, c'est la logique de cohérence curriculaire de type « *top-down* », détaillée plus haut, qui pourrait être remise en question, notamment au niveau de son impact possible sur les pratiques. C'est en tous cas un point qu'il s'agira de thématiser dans cette recherche.

Quoi qu'il en soit, au vu de ces constats, Crahay (2006) propose un axe de questionnement en lien avec notre problématique : « ... il semble légitime de se demander par quel processus cognitif les enseignants composent leurs épreuves d'évaluation et quel paramètre ils prennent en compte pour accomplir cette tâche » (p. 19). Dans ce sens, face à des éléments prescriptifs parfois problématiques, quelles sont les références des enseignants en matière d'évaluation sommative s'ils ne sont pas en mesure de s'appuyer sur un curriculum suffisamment explicite et transparent pour élaborer leurs épreuves ? Cette question renvoie notamment à celle du « *how to do* » dont parle Lenoir (2011, p. 18), au passage à l'acte que le curriculum devrait faciliter du point de vue d'une logique d'alignement curriculaire. Comme le soulignent Demeuse et Strauven (2006) :

il est primordial que le curriculum fournisse aux utilisateurs des directives pour évaluer les apprentissages du public-cible, au moins les apprentissages de base. [...]. Il s'agit de fournir des pistes dans le but de recueillir des informations pertinentes, valides et fidèles autant que possible sur la progression des apprentissages, sur leur degré de maîtrise ou sur la qualité des produits. (p 113).

On observe donc que dans la littérature, les principes définissant les fonctions d'un curriculum relativement à l'évaluation semblent relativement clairs. Mais ces principes se fragilisent une fois confrontés aux prescriptions réelles et aux conceptions des référentiels qui les véhiculent.

Il n'est donc pas aisé, en se fiant à la littérature consultée, de différencier curriculum et référentiel pour qu'ils servent aux enseignants de référence à des fins de construction de cohérence d'alignement curriculaire en évaluation sommative.

Ceci étant, la pratique évaluative se concrétise notamment au travers de l'élaboration de tâches évaluatives. C'est à ce niveau qu'un curriculum, ou un référentiel, peuvent s'incarner et devenir concrets, puisqu'ils sont censés faciliter cette construction en référence à une prescription. Mais qu'en est-il exactement de cette question des tâches évaluatives ?

5.2 Le rôle clé des tâches évaluatives

La littérature sur l'alignement curriculaire accorde une importance particulière aux tâches d'évaluation. Par exemple, Anderson (2002) montre que la correspondance entre le niveau taxonomique et les contenus des tâches évaluatives avec le niveau taxonomique et les contenus des tâches d'apprentissage est à la base d'une cohérence d'alignement curriculaire. Pour lui, l'analyse des tâches d'évaluation doit poser la question suivante : « Avons-nous enseigné ce qui est testé ? ». On voit bien ici le lien de cohérence qu'il met en évidence avec cette interrogation proche de la *praxis*. En effet, la manière avec laquelle les tâches sont construites, tant au niveau de leur forme que de leur fond, donne à voir les apprentissages qu'elles évaluent. Mais, dans la logique de notre modèle, les chercheurs font l'hypothèse que les informations qu'elles délivrent doivent également permettre le recours à des systèmes de pondération et de notation qui mettent en avant la qualité de l'apprentissage. Dans ce sens, les tâches sont censées concrétiser le défi intellectuel que doit relever l'élève, pousser l'enseignant à analyser la complexité de l'apprentissage à mener du point de vue de l'objet et des habiletés, offrir potentiellement des opportunités de travail variées en fonction des contextes et des ressources à disposition (Klenowski & Wyatt-Smith, 2014), tout en offrant des possibilités d'exploiter des critères et des notes qui se réfèrent à des unités d'apprentissages (e. g., Biggs, 1999, 2010). Idéalement référencées aux objectifs prescrits, en lien étroit avec les processus d'enseignement et d'apprentissage offerts aux élèves, les tâches évaluatives matérialisent ce que le curriculum certifie. Dans le même sens, Biggs (2003) spécifie que les tâches évaluatives devraient faire explicitement référence aux apprentissages attendus. Il parle d'ailleurs de tâches inscrites dans des systèmes d'enseignement et d'apprentissage pauvres « *poor system* » lorsque ce n'est pas le cas. Il exemplifie son raisonnement en montrant que souvent, les élèves peuvent très bien obtenir des résultats satisfaisants en restituant des éléments qui n'ont que peu de liens avec les buts de l'enseignement. Autant d'aspects qu'il avait déjà exposés dans un autre écrit plus ancien

(1999), dans lequel il précisait que les tâches d'évaluation doivent rendre possible un jugement sur le niveau atteint par l'élève en rapport avec les objectifs. Mais cette conception de la cohérence demande à être mise en perspective avec la réalité de la pratique, qui montre que régulièrement, les tâches évaluatives sont élaborées en congruence avec d'autres références qu'un curriculum, comme les moyens d'enseignement, les activités de classe ou encore les documents internes à l'établissement. Dans ce sens, elles peuvent aussi être alignées, ou non, à un curriculum réel ou réalisé, voire caché ou implicite (Perrenoud, 1994). Mais est-ce une cohérence acceptable ? Dans quelle mesure ne sommes-nous pas, dans ce dernier cas, face un système trop éloigné du prescrit, et donc possiblement aléatoire ?

Car rappelons-le : la tâche d'évaluation devrait rendre visible la qualité de l'apprentissage évalué en amont, idéalement, en référence aux objectifs du curriculum, et permettre son traitement, en aval, au travers d'un système de pondération en rapport avec l'apprentissage évalué, offrant à son tour la possibilité de construire une note, elle aussi, liée à l'apprentissage. Ce n'est pas rien ! Cet alignement constitue un des enjeux majeurs de cette manière de conceptualiser la cohérence curriculaire de l'évaluation sommative au cœur de laquelle la tâche évaluative joue le rôle de « clé de voûte ». À ce titre, saisie dans ce modèle, elle devrait être « un déclencheur privilégié » (Hadj, 1997, p. 77). Cette métaphore est d'autant plus compréhensible qu'elle image ce que Biggs appelle l'alignement « *bottom-up* » (1999, p. 73) dont nous parlons plus haut.

Les critères d'une tâche évaluative alignée curriculairement vont nous amener à conceptualiser ce que Wiggins (1988) appelle une tâche authentique³⁴. Nous nous y attellerons dans notre cadre conceptuel. Cela étant, notre acception de l'alignement curriculaire allant jusqu'au « traitement » des tâches évaluatives et jusqu'aux pratiques de notation, il nous importe, ici, de convoquer des travaux qui ont documenté la manière avec laquelle les enseignants pondèrent leurs épreuves, et enfin, les modalités qui leur permettent de synthétiser ces pondérations pour produire des notes.

5.3 Une littérature abondante sur la pondération

La pondération est une phase importante de toute démarche d'évaluation sommative. Nous la

³⁴ Ce choix s'est imposé à nous car il met en évidence des caractéristiques des tâches évaluatives en rapport avec notre modèle. Cela ne signifie cependant pas qu'une tâche DOIT répondre aux critères de Wiggins pour être alignée curriculairement.

considérerons ici comme l'action qui vise à mettre en évidence le « poids » de l'apprentissage au travers de l'utilisation de deux outils distincts : les points et les critères. Nous commencerons donc cette partie en abordant la littérature qui a documenté leur utilisation respective. Puis, nous documenterons des travaux ayant pensé la question complexe des critères comme étant les seuls outils pour penser la qualité de l'apprentissage au sein d'une cohérence d'alignement curriculaire jusqu'à la notation. De manière à rester cohérents avec nos cadres théoriques, nous avons volontairement écarté de cette recension les travaux inscrits dans l'approche par compétences, cette dernière étant portée par des fondements conceptuels différents des nôtres.

5.3.1 Pondérer des tâches évaluatives, ou la question de « combien ça vaut ? »

Deux champs de recherches proches fournissent des résultats intéressants sur cette problématique : celui du « *scoring* » ou « *marking* », traduisant littéralement l'acte de pondérer, et celui du « *grading* », relatif à la notation³⁵. Ces résultats émanent principalement de travaux anglo-saxons (d'où ces anglicismes), les plus significatifs sur ces questions selon nous. Premier élément intéressant : assigner de la valeur à des tâches ou des items présents dans une épreuve sommative est un prolongement de toute pratique d'évaluation (McMillan & Nash, 2000). Là, la littérature existante se divise en deux catégories principales : celle qui a investigué la pondération au travers de l'utilisation des points, et celle qui l'a étudiée via l'exploitation de critères.

Des recherches montrent d'abord que c'est la première méthode de pondération, au travers de points, qui est la plus répandue dans les pratiques évaluatives (Marzano, 2002). En revanche, les points s'avèrent unanimement trop proches d'une logique de la mesure pour rendre compte d'une qualité, car ils appellent les totaux, les moyennes, les additions, les comptages, les compensations. Cette logique est dès lors décrite par la recherche en évaluation comme difficilement compatible avec l'apprentissage (Brookhart, 2005). Car, d'une part, la mesure a besoin d'un objet physiquement appréhendable, saisissable dans l'espace/temps, et d'autre part, parce qu'elle nécessite une unité de mesure applicable à cet objet qui soit objectivement définissable. Pour cela, l'objet doit être saisi sous une ou plusieurs dimensions, chacune pouvant être pourvue d'une échelle numérique (Hadji, 2012). À ce titre, Cardinet (1990) est

³⁵ Notons que ces termes définissent des pratiques différentes en fonction de leur origine dans le monde scientifique anglo-saxon. Les chercheurs anglais ont tendance à parler de « *scoring* » pour la pondération et de « *marking* » pour la notation. Leurs collègues américains ou australiens utilisent « *scoring* » pour la pondération et « *grading* » pour la notation. C'est cette seconde option que nous choisissons, au vu des sources retenues.

formel : pour lui, « la mesure n'est pas acceptable en pédagogie » (p. 43), ce que nous approuvons si une des fonctions de l'évaluation sommative est de rendre compte de la qualité de ce qui est évalué dans le but d'opérer des régulations.

Du point de vue de la pratique, Smith, Smith et DeLisi (2001) ont documenté qu'une manière courante de donner des points consiste à déduire des points du maximum prévu en fonction de ce qui sépare la réponse de l'élève de celle qui est attendue. Ce qu'infirme Royce Sadler (2005), lorsqu'il décrit les pratiques de *scoring* comme étant le fait de compter les réponses correctes de l'élève, et que cet acte s'applique essentiellement aux items ou aux tâches, plutôt qu'aux épreuves dans leur entier. Les deux pratiques de comptage de points coexisteraient donc. Mais, fait plus étrange et interpellant, leur exploitation n'est que rarement pensée en référence aux apprentissages (Stiggins, Frisbie & Griswold, 1989, cité par Zhang & Burry-Stock, 2010, p. 325), mais obéit plutôt à une logique comptable. Marzano (2002) précise même que, bien que communes et intuitives, les erreurs induites par cette pratique n'ont pas fait l'objet d'analyses approfondies dans la recherche récente. De son côté, Dubus (2006), affirme que « la tendance spontanée des participants est de compter un point par item réussi, de les additionner de manière à fabriquer une note sur 10, puisqu'il y a 10 items » (p. 63). Il montre qu'un tel système, proche d'une forme de bon sens, est peu cohérent. Il se justifie en avançant qu'une telle manière de procéder, qui offre « 252 façons d'obtenir un score de 5 points sur 10 » (p. 65), devient très discutable, puisqu'elle met l'accent sur une opération mathématique au détriment d'une réflexion sur l'apprentissage. Outre cet arbitraire, et à l'instar de Biggs (2003), il dénonce également les phénomènes de compensation de points, d'univocité de la valeur attribuée aux apprentissages. Ce dernier auteur va encore plus loin lorsqu'il affirme que ces pratiques de pondération traduisent les croyances particulières chez l'enseignant³⁶, à savoir :

- La connaissance est une somme d'unités ;
- Chaque unité en vaut une autre ;
- Elles peuvent s'additionner, se compenser ;
- Peu importe ce qui est correct, pourvu qu'il y en ait assez ;
- Ces unités sont les devises absolues des échelles ;
- L'évaluation est scientifique, précise, objective ;
- La distribution des acquis détermine les résultats.

³⁶ C'est dans un sens générique que Biggs utilise ce terme de croyance, nous le reprenons tel quel.

En lien avec ces croyances, Biggs documente l'existence de procédures spécifiques, assimilables à autant de pratiques évaluatives :

- Une telle manière de pondérer s'effectue en comptant les erreurs ;
- Les tâches évaluatives se compensent ;
- L'évaluation est séparée de l'enseignement ;
- Les conditions d'évaluation sont très bureaucratiques (épreuves communes, échéances, notes punitives, pas de seconde chance) ;
- Les tâches d'évaluation sont décontextualisées ;
- La notation est normative.

Enfin, du point de vue des messages donnés aux élèves, le plus souvent implicites, il mentionne les aspects suivants :

- Toutes les idées se valent ;
- « L'arbre est plus important que le bois »³⁷, à savoir que la forme est plus importante que le fond ;
- La restitution est valorisée ;
- Des facteurs incontrôlables, comme la chance, jouent un rôle important dans l'évaluation ;
- L'évaluation est de la seule responsabilité de l'enseignant ;
- L'auto-évaluation n'a aucune importance.

Comme le montre le chercheur, il est intéressant de noter que tous ces éléments s'inscrivent dans le paradigme de la mesure (De Ketele, 2012), ou *measure model* (Brookhart, 2005) sur lequel nous reviendrons plus loin. En effet, nous y retrouvons la place privilégiée accordée au point comme seul outil standardisé de la pondération. Par ailleurs, l'apprentissage y est considéré comme un objet physiquement appréhendable. Dans cette perspective, l'important pour l'évalué n'est pas d'être performant sur des aspects-clés de l'épreuve, mais d'engranger un maximum de scores, peu importe où. Nous sommes bien loin d'une évaluation de l'apprentissage.

Cela dit, l'usage de points n'est pas frontalement condamné. Et c'est là que certaines sources consultées relèvent la complexité de la pondération. Des travaux ayant approfondi la question comme ceux de Royce Sadler (2005), bien qu'orientés vers une recherche de qualité dans

³⁷ « Trees are more important than wood » (p.4).

l'évaluation des apprentissages, pensent possible le fait d'apparier des points à des critères. Il s'agira donc, dans notre cadre conceptuel, de donner à voir nos choix en matière de système de pondération critérié, pour qu'il soit le plus possible en lien avec une recherche de cohérence d'alignement curriculaire élargi.

Quoi qu'il en soit, la majorité des sources consultées qui ont étudié spécifiquement les pratiques de pondération critiquent avec véhémence le fait de formuler des scores dans des tâches évaluatives uniquement au travers de points. Elles prônent toutes le recours à des critères qualitatifs, dans une perspective critériée, jusqu'à, pour les plus « extrêmes », l'abandon pur et simple des points.

5.3.2 Pondérer des tâches évaluatives, ou la question de « qu'est-ce que cela vaut ? »

La littérature sur l'évaluation critériée est, quant à elle, très abondante. Elle s'est développée dans le monde scientifique autour des années 70, avec la remise en question des perspectives normatives et l'émergence de la pédagogie par objectifs. Elle a pris de l'ampleur et gagné en validité conceptuelle conjointement au développement de l'approche par compétences, celle-ci étant soucieuse de décrire au mieux les ressources exploitées par l'élève dans la situation d'évaluation. Toutefois, ses limites ont été rapidement identifiées, au sein de grilles souvent touffues et peu situées théoriquement : découpage des habiletés cognitives mises en œuvre, absence de contenus, et surtout, problèmes insurmontables à pondérer des critères et à élaborer une note au regard de la grille³⁸. Pourtant, quel que soit le champ auquel appartiennent les chercheurs, la littérature consultée sur le sujet est unanime : l'évaluation critériée est l'unique méthode pour exprimer la qualité de l'apprentissage relativement à sa complexité et à son contenu d'une part (e. g., Jonsson et Svingby, 2007 ; Royce Sadler, 2009), et pour élaborer une note référée à l'apprentissage d'autre part (e. g., Biggs & Tang, 2011 ; Brookhart, 2017 ; Walvoord & Johnson Anderson, 2009). Mais un problème de taille est apparu : une difficulté à formuler une définition consensuelle du critère. Là, nous observons dans les textes consultés qu'il est parfois critère de réalisation, parfois de production, minimal ou de perfectionnement (Dubus, 2006) ; parfois macro, parfois synonyme de l'objectif. Certes, des compromis existent, notamment sur le fait que les critères doivent être « pertinents, indépendants, pondérés et peu nombreux » (Gerard, 2009, p. 77), mais la manière dont cette pertinence, par exemple, est

³⁸ Pour une critique constructive de l'évaluation des compétences, le lecteur se référera à la fin de cette partie et aux ouvrages qui s'y rapportent.

définie, ne fait pas toujours l'unanimité. Deux consensus émergent quand même à ce sujet des sources non exhaustives consultées : l'évaluation critériée permet d'évaluer les apprentissages de l'élève en référence à des objectifs et non en comparaison avec ses pairs, et elle offre à l'élève l'opportunité de se projeter dans l'apprentissage en le rendant visible dès le début d'une séquence (Andrade, 2005 ; Royce Sadler, 2005). Il ressort par ailleurs de ces travaux un constat fort intéressant pour notre problématique : le recours à des critères pousse les enseignants à décrire le niveau de complexité de connaissance attendu au travers de la situation d'évaluation (Marzano, 2002). Cela ne les empêche pas d'apparier ensuite à chaque critère un certain nombre de points, mais ces points seront attribués de façon plus « alignée », puisqu'en rapport avec les apprentissages.

5.3.3 Des enjeux relatifs aux critères analytiques

Cette méthode se rapproche de ce que certains chercheurs appellent une démarche d'« *analytic grading* » (Royce Sadler, 2009, p. 3), fréquemment utilisée dans les pratiques et souvent décrite comme très efficace, où les critères, concrets, débouchent relativement aisément sur une note. Elle constitue une première étape vers une évaluation qualitative car elle offre, selon le chercheur, des pistes de réflexion pour mieux assigner des points en fonction de la valeur de l'apprentissage évalué : « Of particular interest is the convention of assigning weights to items » (p. 265). Ces apports significatifs pour notre problématique sont soulignés par d'autres chercheurs. Par exemple, pour Laveault et Miles (2008), toute approche qui rend les critères d'évaluation plus spécifiques, moins ambigus, donc plus univoques, permet notamment le recours à des échelles descriptives qui participent à la construction d'une note centrée sur l'apprentissage, donc à dominante qualitative. Mais elle est également critiquée par d'autres auteurs, notamment lorsque les critères peuvent se compenser parce qu'ils font chacun l'objet d'une attribution de points qui débouche sur un total (e.g., Biggs, 2003, 2010). Nous sommes donc dans une approche qui réalise un premier pas vers une cohérence d'alignement curriculaire élargi, mais qui comporte tout de même quelques points délicats. Nous y reviendrons dans notre cadre conceptuel.

5.3.4 Des enjeux relatifs aux critères holistiques

L'approche analytique de la pondération est donc un premier pas, mais elle demeure incomplète. Biggs (2003) lui oppose une approche critériée et de notation plus ample, qu'il nomme, avec d'autres, « *holistic grading* » (Isenhour & Kramlich, 2008 ; Royce Sadler, 2009),

au cœur de laquelle intervient le jugement professionnel de l'enseignant. Ce dernier construit ici progressivement des critères plus globaux, renvoyant à la complexité de l'apprentissage, en rapport avec la démarche d'apprentissage de l'élève, référés aux objectifs, englobant tous les aspects évalués. Nous développerons également de manière approfondie cette perspective holistique de la pondération dans notre cadre conceptuel, ainsi que le rôle qu'y joue le jugement professionnel. Toutefois, il est encore ici important de souligner plusieurs de ses points saillants mis en évidence dans la littérature consultée.

Tout d'abord, la méthode ainsi construite vise à permettre des adaptations permanentes, soutenir l'apprentissage et aboutir aussi à une note référée à l'apprentissage. On pourrait alors imaginer, comme Royce Sadler (2009), que la démarche holistique prévaut sur l'analytique. Mais ses constats sont plus nuancés. En effet, la validité des deux dépend de facteurs contextuels (le type d'élèves, les modalités d'enseignement, les contraintes), et elle peut être plus ou moins élevée de part et d'autre. Car il existe chez l'enseignant une idée préexistante d'un « bon » travail qui conditionne l'évaluation, quelle que soit l'approche choisie. Ensuite, un élément déterminant réside également dans le choix et la formulation des critères, qui ont un impact considérable sur le sens donné à l'évaluation : sur quelles bases l'enseignant formule-t-il des critères suffisamment généraux pour mettre en évidence la complexité de l'apprentissage évalué, mais également référés aux contenus pour éviter des biais de jugement ? Enfin, Royce Sadler toujours situe la complexité du jugement professionnel au cœur du processus de traitement des critères à visée de notation. L'évaluation ne se résume pas ici à une opération de mesure, ou à un jugement dont les bases seraient uniquement ancrées sur l'expériences ou le sens commun, mais bien un processus de prise de décision, chaque fois spécifique et référé théoriquement (Klenowski & Wyatt-Smith, 2014). Nous serons amené à approfondir cette question, notamment en convoquant des recherches sur le jugement professionnel. Mais nous observons que la cohérence curriculaire au travers d'une évaluation critériée est bien présente ici.

Dans une revue de littérature conséquente sur le sujet, Jonsson et Svingby (2007) appuient ce constat, et montrent que de nombreux travaux documentent l'impact positif que des critères, qu'ils soient analytiques ou holistiques, ont sur la cohérence de l'alignement curriculaire : clarification du contrat d'évaluation, explicitation des attendus, possibilité de régulations (Denner, Salzman, Harris, 2002). Mais ces constats demandent, selon Jonsson (2014), à être nuancés. En effet, s'il s'avère acquis que tout critère devrait être aligné avec des contenus de la

tâche évaluative, il manque des résultats de recherches sur l'impact que leur nature a réellement sur cette cohérence (Panadero & Jonsson, 2013). Car il ne suffit pas de formuler des critères, même alignés, pour que *de facto*, les performances des élèves s'améliorent. L'enjeu de l'approche critériée, communément admis dans les sources consultées, n'est donc ni plus ni moins ici de rendre visible l'apprentissage. Il y a dès lors une nécessité, d'un point de vue de recherche, à investiguer la nature et le rôle des critères comme des éléments fondamentaux d'une cohérence d'alignement curriculaire élargi. C'est un des enjeux que nous tenterons de relever également dans notre cadre conceptuel.

Un élément ressort enfin de toutes ces recherches : l'approche par critères qualitatifs entretient un rapport étroit avec la validité de l'évaluation sommative. Nous aborderons cette problématique plus loin. Il est pertinent, pour le moment, de terminer l'étude de nos sources relativement à l'alignement curriculaire élargi en traitant des travaux sur la notation.

6 Une notation différente au travers de l'évaluation critériée

L'utilisation de critères d'évaluation constituerait donc une approche pertinente pour une notation à visée qualitative appelée « *holistic grading* » (Royce Sadler, 2009). Cette notation, globale, élaborée à l'aide de critères centrés sur des éléments de contenus significatifs, se distingue d'une notation analytique « *analytic grading* » (p.1), où des critères trop précis participent au découpage du savoir pour finalement se compenser et ne rien évaluer de complexe³⁹. Nous développerons ces apports dans notre cadre conceptuel car ils nous permettent d'appréhender toute la complexité inhérente à un alignement curriculaire élargi jusqu'à la notation. En effet, une littérature importante montre que la démarche pour arriver à une notation holistique est la plus pertinente pour garder une trace de l'apprentissage jusqu'au bilan final. Décrite comme exigeante, elle pose selon nous toute la problématique de la qualité de l'apprentissage et de son traitement dans la note. Et ceci, même si, dans un autre texte, Royce Sadler (2009), soucieux d'arriver à une note « juste », reste à nos yeux très général sur sa fabrication. Cette prudence à faire le pas du système de pondération à la construction de la note se ressent toutefois dans la majorité des travaux cités : évaluer par des critères qualitatifs oblige l'enseignant à se référer à ce qui est attendu au niveau de l'apprentissage, à le caractériser, et à le pondérer. Ces pratiques, qui semblent avoir des implications très importantes sur la note, ne

³⁹ Afin de ne pas surcharger notre propos, nous donnons un exemple de ces démarches de notation contrastées dans notre cadre conceptuel (Partie 3).

peuvent exister sans le recours à une échelle. Dans une perspective de notation holistique, échelle et critères se fondent, puisque chaque échelon de note est caractérisé par des contenus à maîtriser référés aux objectifs. La notation holistique, de notre point de vue, remet donc profondément en question l'échelle traditionnelle.

6.1 Une notation holistique qui questionne l'échelle ordinale traditionnelle

Ce sont les travaux en docimologie qui ont les premiers critiqué cet outil qu'est l'échelle. La référence en la matière reste selon nous De Landsheere (1992). Dans un ouvrage qui a marqué la discipline, il fait état des limites de cet outil, lorsqu'il est considéré d'un point de vue mathématique. Il en identifie quatre faiblesses principales (pp. 146-148) :

- *L'échelle n'a ni zéro ni maximums naturels ou absolus.* Elle commence et finit n'importe où, au choix de l'utilisateur. Donc, même s'ils évaluent un même objectif, du même point de vue, les évaluateurs n'ont jamais des échelles rigoureusement les mêmes. Les différences sont souvent considérables.
- *Les degrés, ou échelons, ne sont pas de la même grandeur à l'intérieur d'une même échelle.* Par exemple, il est impossible de démontrer que la distance séparant la bonne de la très bonne connaissance de l'allemand est la même que la distance séparant la connaissance moyenne de la bonne. Et quid des extrêmes, surtout vers l'excellence ?
- *Des degrés ou échelons correspondant d'une échelle à l'autre pour une même discipline ne sont pas de même grandeur.* Bon en allemand oral pour un enseignant, n'est pas quantitativement égal à bon en allemand oral pour un autre. Rien de peut établir pareille égalité de façon mathématique.
- *Des échelles portant sur des disciplines différentes ne sont pas comparables.* Par exemple, un enseignant qui donnerait plusieurs disciplines dans une même classe serait dans l'impossibilité de se servir des mêmes échelles, les objectifs, l'enseignement, les objets disciplinaires et les situations ne se ressemblant pas.

Ces éléments montrent que les échelles d'évaluation classiques, ordinales, participent essentiellement à quantifier l'apprentissage au travers d'opérations mathématiques. Leur capacité à rendre compte de la qualité de ce qui est évalué est très limitée. Dans ce sens, l'auteur va même jusqu'à affirmer que c'est un instrument mal connu qui vicie une large partie des examens, à tous les niveaux de la scolarité, dans toutes les disciplines. Pourtant, il soulève un paradoxe en montrant conjointement que l'échelle est toutefois incontournable : dans une

perspective non normative, mais qualitative, elle reste le seul outil pour concrétiser un jugement sur une performance pour aboutir à une note, quelle qu'elle soit. Il est dès lors important, toujours selon lui, de respecter certains principes lors de son élaboration :

- L'échelle devrait posséder peu d'échelons, au maximum cinq. Cela semble logique, puisque qualifier un apprentissage ne peut se décliner de manière infinie ;
- Il est nécessaire de décrire chaque échelon en référence aux apprentissages en caractérisant l'objet de l'évaluation au travers de critères. Cette étape permet notamment de critérier le seuil de réussite, si problématique comme nous l'avons notamment évoqué dans notre introduction de thèse, en fonction du contenu évalué et de l'objectif visé. En procédant de la sorte, on s'approche d'échelles holistiques (Biggs ; 2003) ou descriptives (Laveault & Miles, 2008) qui décrivent les différents seuils de réussite d'un point de vue qualitatif.

Ainsi construites, ces échelles sont progressives et mettent des mots sur les acquis et difficultés des élèves. Elles offrent ainsi des éléments de régulation car elles permettent des feedbacks centrés sur les apprentissages. Là, il devient plus aisé de verbaliser à l'élève les éléments de l'apprentissage qu'il ne maîtrise pas en se référant concrètement aux critères.

Toutefois, l'échelle comme outil de construction de la note, peut également être source de problèmes, et ceci, indépendamment de sa nature. Des travaux s'accordent même à dire que le problème de la notation soumise à des échelles ordinales classiques vient moins de la note que de l'échelle (Brookhart, 2011 ; Tierney, 2015), car plus elle possède d'échelons, plus une notation référée aux apprentissages devient difficile et potentiellement biaisée (Jonsson & Svingby, 2007). Mais nous serions tenté de dire que ces problèmes, spécifiquement liés aux caractéristiques mathématiques des échelles traditionnelles, s'estompent dans une démarche holistique. En effet, une telle perspective de notation n'est plus confrontée au dilemme de traduire des apprentissages en termes de quantités ordonnées. En revanche, d'autres problèmes surgissent, comme celui de hiérarchiser la qualité de l'apprentissage⁴⁰.

À la lumière de ces premiers constats sur la notation, il nous incombera donc, dans notre cadre conceptuel, de voir dans quelle mesure une notation holistique est compatible avec des échelles traditionnelles. Nous faisons l'hypothèse que de ce point de vue, les approches théoriques de la

⁴⁰ La manière d'élaborer des échelles qualitatives est toutefois spécifique aux situations d'évaluation et aux apprentissages évalués. Nous montrerons dans nos résultats que ces échelles comportent également des limites et qu'il est dès lors délicat de généraliser leur pertinence.

notation sont à ce point divergentes qu'elles demandent à être conceptuellement tranchées.

6.2 Une notation à visée qualitative : quelques enjeux

Il est maintenant important de documenter à quelles conditions une notation alternative, davantage référée à l'apprentissage, peut être envisagée. Tout d'abord, des recherches anglo-saxonnes ont permis de montrer que si les pratiques de notation s'articulent avec les pratiques d'évaluation, elles se différencient et se pensent spécifiquement (McMillan & Nash, 2000 ; Randall & Engelhard, 2010 ; Walvoord & Johnson Anderson, 2009). En effet, elles documentent que certains facteurs exercent des influences majeures spécifiquement sur les pratiques de notation : les normes prescrites en la matière, la question des seuils de suffisance, le fait de tenir compte de l'effort fourni par l'élève ainsi que des performances supplémentaires démontrées, la distribution des notes, et enfin la question du zéro comme signe d'absence totale d'apprentissage. Par exemple, on apprend que le seuil de suffisance est issu d'un jugement subjectif de l'enseignant, que des notes plus élevées sont données aux élèves qui ont fait preuve de persévérance, ou encore que nombre d'enseignants offrent des occasions d'engranger des points supplémentaires pour augmenter la note (Cross & Frary, 1999). Ces auteurs n'hésitent pas ici à parler de pratiques de notation idiosyncratiques, les enseignants étant toujours en tension entre leurs croyances d'un côté et leurs contraintes de l'autre. Dans de telles perspectives, comment penser une notation référée à l'apprentissage ? En effet, on observe ici une difficulté chez les praticiens à justifier leurs pratiques de notation au regard de pratiques d'évaluation en lien avec des contenus et objectifs, même si l'on constate dans ces recherches qu'ils sont unanimes à reconnaître l'interdépendance entre notation, évaluation, enseignement et apprentissage, et cela, même si les enseignants incluent dans leurs constructions de notes des éléments indépendants d'apprentissages cognitifs tels que la motivation, l'engagement ou l'effort (Cross & Frary, 1999 ; Martinez, Stecher & Borko, 2009). Sur ce point, la littérature consultée semble converger vers l'idée qu'évaluer d'un point de vue sommatif des éléments n'ayant souvent pas fait l'objet d'un enseignement, et de surcroît délicats à caractériser, induit des biais dans l'élaboration des notes (e. g., Zhang & Burry-Stock, 2003), et donc converge vers un désalignement⁴¹ curriculaire. Il devient dès lors important, d'un point de vue de recherche, de questionner à quelles conditions la notation peut se centrer sur les apprentissages réalisés et en rendre compte. En d'autres termes, à la lumière de ces apports, on se demandera

⁴¹ Dans notre travail, et fidèlement à la littérature consultée, nous considérerons désalignement et rupture d'alignement comme des synonymes traduisant une perte de cohérence de l'évaluation sommative.

quels raisonnements vont permettre de penser une notation rendant visibles les apprentissages, et à l'aide de quels outils cette visibilité peut devenir envisageable. Cette réflexion demande à problématiser la notation comme une dimension qualitative du processus d'évaluation sommative.

6.3 Des notes pour une évaluation sommative au service des apprentissages

Les travaux qui ont abordé cette problématique d'une notation en référence aux apprentissages - c'est-à-dire pouvant rendre compte des apprentissages réussis mais aussi insuffisants dans une perspective de régulation - proposent tous, en amont, de passer par un travail d'explicitation des exigences attendues. La question clé revient ici à se demander ce que doivent dire les notes (Brookhart, 2011). Ce raisonnement implique, dans un premier temps, une référence à des objectifs clairs, univoques et compréhensibles (Klenowski & Wyatt-Smith, 2014), reliés à des tâches évaluatives en rapport, travail impossible à réaliser sans une expertise des contenus à enseigner (Bateman et al., 2009). Puis, dans un deuxième temps, l'enseignant aura recours à des critères d'évaluation centrés sur des contenus et en cohérence avec les tâches et objectifs évalués. Comme nous l'avons déjà évoqué, cette évaluation critériée est, selon toutes les sources que nous avons consultées, la méthode incontournable pour une notation référencée aux apprentissages.

6.4 Des constats fondamentaux sur une notation référée aux apprentissages

Cette partie de notre revue de littérature sur la notation qualitative nous permet d'établir deux constats importants. Le premier est qu'il est possible de considérer qu'une évaluation sommative référée aux apprentissages peut aboutir à une note qui donne des informations qualitatives sur ces derniers (e.g., Pellegrino, 2006 ; Walvoord & Johnson Anderson, 2009). Le second constat, directement en lien, nous fait affirmer qu'il est possible d'envisager une approche englobante de la question de la cohérence de l'évaluation, et donc de sa validité. L'étude des liens de cohérence existant entre la référence au prescrit, les tâches évaluatives, leur pondération et la notation ouvre des possibles, tant d'un point de vue de recherche que de développement des pratiques évaluatives sommatives. Nous touchons ici au cœur de notre orientation théorique : l'alignement curriculaire élargi.

7 Des recherches pour une note alignée curriculairement

Les sources auxquelles nous nous référons ici construisent leur raisonnement sur la notation à partir de l'évaluation critériée. Aucune recherche n'a, à notre connaissance, montré que l'unique utilisation de points constitue un outil qui qualifie l'apprentissage dans une recherche de cohérence d'alignement curriculaire élargi. Certes, nous l'avons vu, nous trouvons des auteurs qui considèrent l'appariement des points et des critères envisageables, mais ils ne problématisent pas cette association dans une recherche de cohérence d'évaluation (e. g., Royce Sadler, 2009). Reste que d'un point de vue de pratique évaluative, il existe un passage entre réaliser des grilles de critères et les « transformer » en une note, ce que Hadji considère comme une transaction (2016). Walvoord et Johnson Anderson (2009) abondent dans ce sens, lorsqu'ils affirment : « rubric scoring and grading are not the same thing » (p. 44). En cela, ils analysent les conditions pour qu'il y ait un lien de cohérence entre les critères et leur exploitation dans une échelle. Avec ce raisonnement, ils vont selon nous plus loin théoriquement que Dubus (2006), ici inscrit dans des approches de mesure, qui affirme, sans justification, que « chaque critère gouverne l'attribution d'une certaine quantité de points » (p. 96).

7.1 Une note sans mesure ?

Nous l'avons déjà évoqué, pour certains chercheurs, la note, de fait, est une mesure. Dans ce sens, elle ne peut pas apporter des éléments qualitatifs sur l'apprentissage évalué. Suite à l'analyse des nombreuses sources citées, nous soutenons la thèse inverse, à savoir que non seulement la note est susceptible d'apporter des éléments qualitatifs sur les apprentissages des élèves, mais qu'elle devrait tendre vers cette finalité. À notre sens, c'est parce que la note a longtemps été assimilée uniquement à une mesure qu'elle est encore aujourd'hui stigmatisée dans certaines communautés de chercheurs et de mouvements éducatifs dont nous parlons plus haut, ainsi qu'au sein de groupes d'enseignants. Toutefois selon certains chercheurs ayant investigué cette problématique au sein du modèle de l'alignement curriculaire (e. g., Biggs, 2003 ; Biggs & Tang, 2011 ; Martone & Sireci, 2009), il apparaît, ici aussi, que le problème se situe moins au niveau de l'existence de la note que de ses processus de construction (Hadji, 2017). Nous retrouvons cette problématique de la construction déjà évoquée. Mais il est intéressant de constater que pour la traiter, les recherches consultées issues du champ des curricula se réfèrent aux outils créés et utilisés pour les produire.

7.2 Quelle échelle adopter ?

Dans la littérature consultée, deux caractéristiques de ces outils apparaissent déterminantes : le nombre d'échelons et le barème. Nous discutons ci-après leurs caractéristiques au regard de notre modèle de l'alignement curriculaire élargi.

A propos du nombre d'échelons, Walvoord et Johnson Anderson (2009) considèrent qu'il doit être en rapport avec l'objet et les objectifs évalués, et ainsi répondre à la question « how many distinctions do you need to make ? » (p. 47). Nous retrouvons ici le lien assez explicite avec une recherche de cohérence dans l'alignement curriculaire élargi, dans la mesure où l'on interroge la déclinaison possible de valeurs dans l'évaluation de l'apprentissage attendu : quel apprentissage désigne la suffisance, l'insuffisance, l'excellence ? Lorsque le nombre d'échelons est imposé⁴², ou quand l'échelle est issue d'une banque de données extérieure ou de directives du législateur, ou encore lorsqu'elle émane d'habitudes non questionnées, elle entre en conflit avec la logique de cohérence curriculaire construite en amont et la question devient alors très problématique. Le postulat affirmant que le problème est moins la note que l'échelle (Royce Sadler, 2009) perdure ici aussi, au cœur de notre modèle, et de manière encore plus prégnante. Comme le dit Biggs (2003), avoir recours à des échelles standardisées provoque, de fait, une rupture d'alignement curriculaire, quel que soit le raisonnement réalisé sur la construction de l'épreuve sommative, et ceci encore davantage si la perspective s'avère normative, puisque l'échelle est construite indépendamment de l'apprentissage et des conditions qui l'ont permis. Une littérature nous montre donc que la construction de l'échelle, de ses barèmes, et de la description de ceux-ci à l'aide de critères qualitatifs référés à des objets et à des objectifs, fait partie intégrante de la cohérence curriculaire d'une évaluation sommative, au même titre que la construction des tâches évaluatives ou de leur pondération. À notre connaissance, peu de recherches ont abordé cette problématique, notamment du point de vue des ruptures d'alignement qui peuvent exister au sein de chaque relation.

7.3 Quel barème élaborer dans ce cas ?

Approfondir cette problématique nous invite donc également à aborder la question du barème. Une des rares définitions que nous en avons trouvé est celle de Dubus (2006) : « une liste d'éléments distincts (munis de *critères*) auxquels sont attachés des points qui seront totalisés

⁴² Peut-être est-il utile de rappeler que dans la scolarité obligatoire du canton de Vaud, les échelles, imposées, sont constituées de onze échelons, les notes allant de 1 à 6, avec demi-notes.

pour fournir la *note* ou *score* total » (p. 264). Nous retrouvons ici deux éléments problématiques que nous avons déjà mis en discussion : une association entre critères et points au service d'une construction de note avec tous les problèmes de cohérence que cela comporte ; et une absence de référence aux apprentissages évalués. Cet état de fait est interpellant d'un point de vue de recherche, puisque des travaux ont abordé la question d'une note à visée qualitative, mais aussi au regard des pratiques réelles d'évaluation sommative, qui ne peuvent exister sans barème ! Ceci dit, nous nous accorderons tout de même sur le fait que tout barème vise à réduire les jugements exprimés de manière disjointe sur les critères en un résultat global, au travers d'un processus de hiérarchisation distinguant le bon du moins bon. En cela, un barème est une sorte de « garde-fous » (Dubus, 2006, p. 95) pour aller vers une forme d'objectivation du résultat.

Assimilé à un outil, le barème a également été étudié au sein de la docimologie critique. Cette dernière, influencée par la psychométrie, a montré les conséquences arbitraires que le barème a sur l'évaluation des élèves (Mottier Lopez, 2015), lorsqu'il est exploité comme outil de mesure des apprentissages. Là, il est assimilé par certains comme outil de pouvoir (Desclaux, 2014). Mais il sert également la limitation des aléas pour améliorer la notation (Merle, 2012a). En effet, le barème constitue pour certains chercheurs le lieu où la qualité de l'apprentissage est rendue visible, pour que la notation soit aussi en cohérence avec l'apprentissage. En d'autres termes, l'enseignant devrait savoir maîtriser les techniques de construction de barèmes (Tessaro, 2013), en référence aux objets évalués et à leur pondération au sein des épreuves. Cette tâche est d'une grande complexité, comme le montrent de nombreuses études (e. g., Reeves, 2004 ; Walvoord & Johnson Anderson, 2009), car, encore une fois, elle demande d'une part l'expertise des contenus évalués (en termes d'apprentissage) et la connaissance d'éléments relatifs à la notation.

On le voit, le barème peine à être défini dans la littérature, sauf quand son épistémologie est très orientée sur la mesure (De Ketele & Roegiers, 2009 ; Grégoire & Laveault, 2002 ; Morissette, 1997). Ce constat nous amène à choisir une autre conceptualisation de l'échelle - et du barème qui y est associé - en privilégiant des recherches qui ont pensé la constitution des outils d'évaluation sommative non pas en fonction de leur capacité à différencier l'ordre des performances des évalués, mais plutôt de leur capacité à mettre en évidence la qualité produite.

8 Au centre : une prise d'informations référées à un modèle théorique

Au-delà des convergences et tensions que les travaux consultés mettent en évidence

relativement aux objets spécifiques qu'ils abordent, nous constatons qu'ils se rejoignent sur deux points. Le premier considère qu'une évaluation sommative à visée qualitative ne peut se réaliser, dans une perspective de recherche de cohérence d'alignement curriculaire élargi, sans un recours au jugement professionnel de l'enseignant (e. g., Biggs, 2003). Et ce n'est qu'en sélectionnant un certain nombre d'informations pertinentes référées à ce modèle que ce jugement peut se déployer. Le deuxième point souligne que ce jugement en évaluation s'inscrit dans une recherche permanente d'indices relatifs à la qualité de l'apprentissage évalué.

8.1 Le jugement professionnel de l'enseignant : un geste omniprésent

Le concept de jugement professionnel est fréquemment utilisé mais peu empiriquement éprouvé. Sa théorisation est assez récente. Réalisés avec de nombreux collègues, les travaux sur le sujet de Mottier Lopez s'accordent sur le fait que le jugement professionnel de l'enseignant intervient dans l'ensemble de ses activités professionnelles : dans le choix et l'agencement des situations didactiques et pédagogiques, des méthodes de gestion de classe, des modalités de collaboration, des registres de communication avec les parents, et évidemment dans toutes les démarches d'évaluation d'apprentissage des élèves (Allal, 2009). Le jugement est donc ici entendu comme un « acte de discernement », et non comme un jugement normé (Mottier Lopez & Tessaro, 2016). Relativement à l'évaluation sommative, certaines études, menées par Mottier Lopez (2008), Allal et Mottier Lopez (2008) ou encore Mottier Lopez et Cattafi (2008) mettent en évidence, par exemple, que le jugement professionnel en évaluation n'est pas un acte mécanique, mais qu'il s'inscrit dans un réseau de communication et d'échanges entre les différents partenaires concernés. Par ailleurs, sur quelques étapes clés de la démarche sommative, l'analyse de certaines données « montre que des significations sont attribuées aux notes, en plus de leur définition en termes de degré d'atteinte des objectifs, et que ces significations jouent un rôle déterminant en cas d'incertitude, telle que l'hésitation entre deux notes » (Mottier Lopez & Allal, 2008, p. 476).

En évaluation sommative, le jugement professionnel intervient donc dans toutes les étapes de son élaboration, que ce soit au niveau du choix des outils, de la pondération des critères, de la formulation des appréciations ou encore de l'exploitation des résultats. Nous faisons dès lors l'hypothèse qu'il existe un jugement professionnel en évaluation spécifique à des situations d'évaluation sommative. Ce jugement évaluatif impliquerait « la mise en relation du singulier (tout ce que l'évaluateur sait de l'individu concerné) et du général (savoirs disciplinaires formels, normes et règles institutionnelles) afin de décider d'une action à la fois bien adaptée

et bien contextualisée » (Allal, 2009, p. 190). Mais une telle conception rend-elle compte des articulations entre les processus d'enseignement, d'apprentissage et d'évaluation sommative ? Et qu'entend-on par général ? Du point de vue de l'évaluation sommative, cette acception du général est-elle ici suffisante pour exprimer cette nécessité de prendre en compte des informations valides dans tout processus de jugement référencé aux apprentissages ?

8.2 Le jugement professionnel au service d'une cohérence d'alignement curriculaire

Dans notre perspective, c'est au travers de notre modèle que le jugement professionnel devrait donc s'inscrire, comme un processus décisionnel transversal et systémique au service de décisions inscrites dans une recherche de cohérence curriculaire.

La littérature consultée dans ce champ montre que toute pratique d'évaluation s'appuie sur un jugement professionnel, lui-même inscrit dans une recherche de cohérence curriculaire, du choix des objectifs d'apprentissage à évaluer, jusqu'à la construction de la note. Le jugement professionnel de l'enseignant, dans une perspective de développement d'évaluation sommative qualitative, intervient donc au cœur d'une recherche permanente de cohérence d'alignement curriculaire élargi. Dès lors, d'un point de vue de recherche et de formation, il ne s'agirait pas seulement d'étudier par exemple comment les enseignants s'y prennent pour juger de la pertinence des réponses des élèves, ou de voir quelles informations ils sélectionnent pour décider d'une note, mais de comprendre quels invariants participent à la formulation du jugement lorsqu'il s'agit de mettre ces divers niveaux de la pratique évaluative en lien. Cette démarche pourrait favoriser ce que Laveault (2008) appelle un *cercle vertueux*, puisque « le jugement professionnel contribue à améliorer les pratiques d'évaluation et ces pratiques permettent en retour une meilleure réflexion sur les résultats, sur le suivi de l'élève et le support à lui apporter » (p. 493). L'amélioration des pratiques évaluatives sommatives se réaliserait donc au travers d'un jugement professionnel - considéré comme une véritable capacité à construire une intelligibilité des phénomènes d'évaluation en situation (Mottier Lopez & Allal, 2008) - au service de l'élaboration d'une cohérence d'alignement curriculaire élargi et référé à celui-ci. À notre connaissance, aucune recherche n'a investigué cette problématique à ce jour dans ce modèle. C'est la raison pour laquelle, dans notre cadre conceptuel, nous tenterons de définir un jugement professionnel « holistique ». Car c'est le seul, à nos yeux et conformément à la littérature sur l'alignement curriculaire et à l'approche qualitative qu'il sous-tend, qui est en congruence avec notre modèle, notamment pour assurer une validité de l'évaluation sommative.

9 Pourquoi ne parler que de la validité de l'évaluation ?

La question est volontairement ici provocatrice. La littérature sur l'évaluation des compétences s'est attelée davantage à une définition plus large de la qualité de l'évaluation sommative, en insistant sur le fait qu'en tant que recueil d'informations, elle devait répondre à des critères de fiabilité, de pertinence et de validité. Ces trois concepts, bien connus dans les approches de la mesure et de l'évaluation, ont été redéfinis par De Ketele et Gérard (2005) au cœur de l'approche par compétences. Ils postulent qu'ils sont classiques lorsqu'il s'agit de questionner cette qualité. Pour eux, la pertinence renvoie à la cohérence de l'épreuve au regard des objectifs visés ; la fiabilité est le degré de confiance accordé aux résultats même s'ils sont récoltés à un autre moment, avec un autre outil et par un autre évaluateur ; et la validité renvoie « au degré d'adéquation entre ce que l'on déclare faire et ce que l'on fait réellement, entre ce que l'outil mesure et ce qu'il prétend mesurer » (p. 3).

Ces définitions posent selon nous plusieurs problèmes. Tout d'abord, leur niveau de généralité ne permet pas de considérer l'impact des objets de savoir évalués ni la spécificité des tâches évaluatives sur l'évaluation sommative. Ensuite, au niveau de la pertinence, qu'entend-on par « cohérence de l'épreuve » ? De quelle cohérence parle-t-on, du point de vue de l'apprentissage ? Peut-on par ailleurs généraliser cette question de l'épreuve à toutes les disciplines, comme s'il n'en existait qu'une seule déclinaison ? Par ailleurs, quel est le sens de traiter de la fiabilité lorsque l'on postule que l'évaluation est, de fait, une pratique pédagogique, sociale et institutionnelle (Mottier Lopez, 2015), et donc toujours représentative d'un contexte rarement reproductible, surtout si l'on se soucie des apprentissages des élèves et non de normes extérieures ou de la standardisation des outils ? Comme le disent Cardinet et Laveault (2001) : « Force est donc de conclure que la détermination du score “vrai” est impossible, et que toute mesure est une fonction à deux variables, où l'examineur (ou bien la méthode d'évaluation choisie) a autant d'influence sur le résultat que la compétence de l'examiné » (p. 19). On pourrait questionner l'usage de la mesure dans cette affirmation, et dans le même sens, se demander pourquoi traiter de la validité à l'aune de cette dernière ? Enfin, la validité n'est-elle pas, justement, l'expression d'une évaluation dont le but serait essentiellement... d'évaluer⁴³, et

⁴³ Ce qui rejoindrait la caractérisation de l'évaluation réalisée par Allal (2008), qui dit : « L'évaluation est constituée d'une série de démarches : 1/ la définition de l'objet d'évaluation; 2/ la récolte d'informations concernant les conduites des apprenants en rapport avec l'objet choisi; 3/ l'interprétation des informations recueillies; 4/ la prise de décisions et la communication des appréciations à autrui (apprenants, parents, collègues, cadres, ...) ».

non de mesurer ? Bref, d'un point de vue théorique, nous ne partageons pas cette acception d'une cohérence d'évaluation sommative en trois critères, car il nous semble très discutable de dissocier pertinence et validité. En effet, dans notre modèle théorique où l'objet et les objectifs évalués occupent une fonction centrale, la validité rend compte également de ce que ces auteurs appellent la pertinence. Il nous faut donc nous orienter vers d'autres travaux pour justifier ce choix.

Ce sont les recherches de Black et al. (2010) sur la validité des pratiques évaluatives qui nous ont apporté le plus d'éléments significatifs. Leur modèle théorique fait clairement référence à une conception de la validité à l'aune de la cohérence de l'alignement curriculaire. Pour eux :

« The model of the determinants of validity in terms of a linked chain of the relevant components was a useful heuristic, for it served to direct attention to such matters as the tasks used, their administration and scoring, the aggregation of scores, the evaluations and judgements which followed, and the actual impact of the results » (p. 217).

On observe donc déjà ici que la pertinence - à savoir le rapport entre les éléments de contenu de l'évaluation et les objectifs évalués - est au centre, et qu'elle n'a dès lors pas à être définie en amont. Notons au passage que cette conception est particulièrement présente dans les recherches qui ont documenté les pratiques d'évaluation sommative critériées (e. g., Jonsson & Svingby, 2007 ; Moskal & Leydens, 2000 ; Reddy & Andrade, 2010), ce qui n'est pas dû au hasard. Comme nous l'avons dit au point traitant spécifiquement de la littérature en rapport, les recherches dans ce champ soulignent que le recours aux critères rend notamment visibles les caractéristiques des objets disciplinaires évalués (Royce Sadler, 2009). De ce fait, cette méthode d'évaluation participe à construire la chaîne - « *linked chain* » - dont parlent Black et ses collègues, une chaîne dont les maillons devraient être les apprentissages.

L'évaluation critériée infléchit positivement cette validité, car elle offre la possibilité, en amont, de mettre cette dernière en rapport avec les objectifs visés, et en aval, avec la pondération et ensuite la notation. Un raisonnement que nous retrouvons chez d'autres chercheurs qui définissent la validité comme le seul indicateur de qualité des informations récoltées (e. g., Brookhart, 2005 ; Biggs, 2010). Penser l'alignement curriculaire élargi en ces termes met alors en relation les questions de validité et de fiabilité, qui se retrouvent dans un rapport dialectique : « Reliability and validity are interconnected » (Royce Sadler, 2009, p. 162). À l'instar de certains auteurs, nous faisons l'hypothèse que la fiabilité doit ici être comprise comme le fruit

d'une démarche visant à construire de la validité à l'aune d'une cohérence d'alignement curriculaire, plutôt que comme un processus visant une généralisabilité des résultats. Comme l'affirment Klenowski et Wyatt-Smith (2010) : « So, while emphasizing the importance of reliability, without validity, reliability is viewed as trivial » (p. 52).

Isoler la question de la fiabilité n'aurait donc que peu de sens lorsqu'elle est saisie dans une problématique comme la nôtre. Et poser la question de la pertinence en amont guère davantage. En revanche, cette approche conceptualise plus précisément la question de la validité de conséquence, qui « considère comme juste et appropriée toute forme d'évaluation qui exerce des “conséquences” positives sur les personnes évaluées » (Laveault, 2004, p. 253). De niveau assez général, cette définition prend du sens dans notre modèle théorique, puisque les conséquences positives d'une évaluation cohérente s'incarnent dans des feedbacks rendus possibles grâce à une recherche de qualité dans les bilans sommatifs. Ce qui nous permet de postuler, à l'instar de certains chercheurs comme Bateman et ses collègues (2009), qu'une évaluation sommative alignée curriculairement - et dès lors valide - produit des jugements plus proches des apprentissages des élèves. Par ailleurs, nous serions tenté d'aller plus loin encore, en affirmant, comme le suggèrent Cardinet et Laveault (2001), qu'une telle évaluation participerait au développement d'une validité systémique, puisqu'elle « entraîne une transformation des pratiques » (p. 28). Nous retrouvons donc ici un lien assez clair avec la qualité de l'évaluation qui, à nos yeux, est en congruence avec notre ancrage théorique, mais qui nous obligera, dans notre partie sur le sujet, à conceptualiser le rapport entre cohérence de l'alignement curriculaire élargi et validité de l'évaluation.

On le voit, l'acception de la validité dépend fortement des épistémologies et cadres théoriques qui infléchissent les différents modèles d'intelligibilité de l'évaluation sommative, notamment d'un point de vue paradigmatique. Et c'est justement parce que cette question des paradigmes, centrale, est parfois implicite dans les recherches sur l'évaluation, qu'il nous semble maintenant important de nous y pencher.

10 Les paradigmes d'évaluation : des systèmes de production de jugements spécifiques

Nous l'avons montré, la majeure partie des recherches portant sur l'évaluation, la pondération et la notation font, de près ou de loin, référence à deux dimensions omniprésentes dans toute pratique évaluative : la mesure et le jugement. Ces deux dimensions sont les caractéristiques de systèmes de productions de jugements - que l'on nomme paradigmes - très contrastés.

Le paradigme, au sens de Kuhn (1970/2008), désigne un ensemble cohérent de postulats regroupant « des lois, des théories, des applications et des dispositifs expérimentaux qui doit fournir des modèles qui donnent naissance à des traditions particulières de recherche » (p. 30). Toutefois, le concept pose problème. Vial (2008) affirme que Kuhn en a fait un véritable passe-partout, puisque dans son ouvrage coexistent pas moins de vingt-quatre sens de ce mot ! Un constat qui met en évidence ici la complexité de la question, et la difficulté, voire l'impossibilité à « capturer » ce concept de paradigme dans des définitions simples et consensuelles. C'est peut-être une des raisons pour lesquelles Mottier Lopez (2015) questionne la possibilité de définir un seul paradigme d'évaluation intégrateur. Elle montre qu'à ce niveau, « les avis des auteurs divergent et que le débat est ouvert » (p. 38). Toujours est-il que ces divergences aboutissent à un constat : bon nombre de travaux consultés font comme si le paradigme dans lequel ils s'inscrivent allait de soi. Malgré les problèmes théoriques que soulève le concept, cet état de fait nous semble très délicat, surtout lorsqu'il s'agit d'analyser les pratiques évaluatives et les outils sur lesquels elles s'appuient. En effet, d'un point de vue d'évaluation sommative, il est fondamental d'avoir conscience des conditions auxquelles les lois, les théories, les dispositifs ou encore les outils peuvent, ou non, cohabiter, afin d'assurer une cohérence d'alignement curriculaire élargi dans l'évaluation. Il nous paraît également important de comprendre l'impact que ces éléments ont sur la production des jugements évaluatifs des enseignants. L'enjeu est ici central, notamment parce que c'est au travers de telles clarifications que la recherche pourra aboutir à des modèles éprouvés. Allal (2012), à ce sujet, est catégorique : « dans chacun des champs de l'évaluation, l'activité évaluative est orientée par des paradigmes de référence qui répondent à des exigences de la pratique très différentes, des exigences que je considère *in fine* comme étant inconciliables » (p. 182).

D'un point de vue théorique, De Ketele (2012) est plus modéré, notamment dans sa définition de quatre paradigmes, avec autant d'épistémologies⁴⁴. Deux d'entre eux nous intéressent plus particulièrement dans le cadre de notre recherche : le paradigme de *l'évaluation mesure* et celui de *l'évaluation de la maîtrise des apprentissages*, car ils émergent des pratiques d'évaluation sommative. Dans le premier, les jugements sont produits en référence à des normes et outils

⁴⁴ Notons que l'auteur définissait dans les années 2000 douze paradigmes, et tente aujourd'hui une conceptualisation d'un seul paradigme intégrateur, comme Figari (2012), qui note qu'il existerait « un début de consensus d'un nombre croissant de chercheurs autour de paradigmes stabilisés, [...] autour de quelques "unifications" ou "intégrations" » (p. 20). Toutefois, dans une perspective d'étude des pratiques réelles, c'est la théorisation en quatre paradigmes que nous retenons ici, car nous la trouvons plus adaptée à une investigation de la *praxis*.

extérieurs, aboutissent souvent à des comparaisons entre évalués, se réfèrent aux performances cibles, donnent une importance aux distributions des scores. Le second induit des jugements réalisés en référence aux objectifs, met au centre l'efficacité des apprentissages et la régulation au travers d'un jugement professionnel situé. Caractérisés de la sorte, ces deux paradigmes vont notamment nous aider à comprendre comment les pratiques évaluatives sommatives peuvent gagner en cohérence curriculaire lorsqu'elles sont prises dans des systèmes de production de jugements aussi divergents. Ce qui devrait être, de notre point de vue et selon une perspective pragmatique assumée, la finalité ultime d'un tel concept. En effet, comme le dit Figari (2012) :

« Le paradigme permet de dire *l'épistémé* (l'évaluation appartenant au questionnement scientifique et se réclamant de postulats, de règles et de méthodes qui engagent une communauté de chercheurs) tout en affirmant, en même temps, la doxa (l'évaluation demeurant une pratique établie de jugement, demandant seulement à être instrumentée) » (p. 222).

Les enjeux de pratique évaluative que soulèvent les questions de validité d'évaluation et de confrontation des paradigmes sont donc nombreux. D'un point de vue de recherche et de formation, il devient nécessaire, selon nous, de les traiter conjointement.

11 Quelques points de convergence pour une épistémologie de l'évaluation

Comme le dit Figari (2012), « on trouve, dans l'ensemble de la littérature sur l'évaluation, des positions tantôt a-épistémologiques, tantôt pré-épistémologiques, tantôt ouvertement épistémologiques » (p. 213). Ces dernières sont définies comme « interrogeant les pratiques en termes de dimensions scientifiques, de modélisation et de production de connaissance et/ou de rapport au savoir » (Ibid., p. 213).

Il nous semble ici avoir montré, au travers de l'analyse de nombreuses sources permettant de déboucher sur nos choix théoriques, que notre position sur l'évaluation est épistémologique, dans la mesure où elle tente de donner à voir des phénomènes évaluatifs issus de la pratique et d'en envisager une conceptualisation en se référant à un modèle théorique précis.

12 Une revue volontairement orientée

Dans cette partie, notre but, en nous appuyant sur des travaux existants, est de poser les bases d'un débat scientifique constructif propre à notre problématique, afin d'aboutir à la compréhension du choix de notre modèle théorique de référence. Nous sommes conscient que

ce dernier - l'alignement curriculaire élargi - est relativement original dans la recherche francophone qui, à notre connaissance, ne l'a jamais conceptualisé de cette manière. Depuis les années 2000, c'est un autre modèle qui monopolise l'attention de la recherche en évaluation : l'approche par compétences. Nous avons souhaité nous en distancer, pour des questions d'incompatibilités théoriques et épistémologiques avec notre problématique.

Un des premiers à avoir critiqué cette approche à notre connaissance est Bain (2002) dans un texte où il montre, au travers d'exemples finement analysés, combien évaluer des compétences en classe est peu valide, notamment parce que la conception de la complexité qui y est associée ne s'accorde pas avec la forme scolaire, d'un point de vue de situation évaluative, mais aussi de notation. D'autres, à l'instar d'Allal (2002), ont montré combien cette approche est inappropriée dans une évaluation scolaire traditionnelle et donc conditionne les modalités d'évaluation sommative. Elle affirme par exemple que l'évaluation de compétences « implique le remplacement des tests classiques composés d'une série d'items indépendants par des situations d'évaluation exigeant des productions effectuées en plusieurs étapes à l'aide d'outils et/ou de formes d'interactions sociales caractéristiques du domaine en question » (p. 90). Est-ce réaliste ?

Parallèlement, et c'est à nos yeux un point très interpellant, des auteurs ont émis des doutes sur l'incomplétude théorique de la notion de mobilisation, au cœur de l'approche (e. g., Carrette, 2007 ; Crahay, 2006), ou sur les notions de ressources dont les définitions vont jusqu'à parfois être contradictoires (Crahay, 2006). Enfin, la compétence renvoyant à la complexité, nous pourrions nous demander quelle acception théorique de cette dernière est conciliable en situation d'évaluation scolaire. Où commence-t-elle ? Comment se traduit-elle ? À l'aide de quels apports théoriques est-on en mesure de l'appréhender pour une évaluation sommative au service des apprentissages ? Là aussi, les tentatives de réponses divergent.

Quant à la notation saisie dans cette approche, elle n'a été à notre connaissance que très peu étudiée, ou théorisée. N'y aurait-il donc pas urgence à éprouver une démarche originale de théorisation des pratiques évaluatives sommatives ?

Nous avons tenté de montrer, dans cette partie, que c'est notamment à ces nombreuses questions et incomplétudes que notre recherche tente de répondre. Cette revue de littérature nous a permis de déterminer nos choix conceptuels que nous exposons et argumentons dans les deux parties suivantes. À ce titre, certains éléments que nous abordons ici y seront à nouveau traités, mais

de manière plus approfondie et dans la perspective de définir et justifier nos choix théoriques.

Partie 3 - Cadre conceptuel relatif à l'étude des pratiques évaluatives des enseignants

1 Introduction

Un principe semble partagé dans la littérature lorsqu'il s'agit de savoir si la pratique évaluative repose sur des fondements théoriques (Figari, 2012) : une théorie de l'évaluation englobe un ensemble de concepts, d'approches de la connaissance ainsi que des modèles compréhensifs (Dubois & Marceau, 2005), et ce sont justement ces modèles compréhensifs de l'évaluation qui aident les acteurs à justifier leurs choix et leurs pratiques. En suivant Figari (2012, p. 219), nous citerons de manière non exhaustive les modèles *conceptuel* (permettant de structurer la pratique évaluative), *explicatif* (offrant la possibilité de révéler les phénomènes évaluatifs), *interactionniste* (permettant de comprendre la dynamique de la pratique évaluative en fonction du contexte) ou encore *prescriptif* (offrant des registres de pratiques évaluatives efficaces) qui ont marqué - et marquent encore aujourd'hui - l'épistémologie de l'évaluation. Toutefois, l'acception de la notion de modèle est très vaste dans la littérature, met en évidence une forme d'éclatement (Figari, 2006 ; Dortier, 2006), jusque dans ses tentatives de typologisation. C'est la raison pour laquelle, au regard du développement important de l'évaluation en éducation, le modèle⁴⁵, malgré son omniprésence, mérite aujourd'hui d'être interrogé dans une perspective épistémologique (Mottier Lopez & Figari, 2012).

Dans le cadre de notre recherche, nous le définirons comme « une représentation conceptuelle qui permet de rendre intelligible une situation, un phénomène ou un processus » (Le Moigne, 1999 cité par Mottier Lopez & Figari, 2012, p. 14). Dans ce sens, nous considérerons que le modèle implique une relation entre une théorie et le pragmatisme de la pratique évaluative. De ce fait, et d'un point de vue épistémologique, son exploitation éclaire soit une posture déductive, soit une posture inductive, soit encore une posture abductive (Anadon & Guillemette, 2007 ; Mottier Lopez & Dechamboux, 2017). La première consiste à mobiliser un modèle théorique connu et validé, comme grille de lecture d'un phénomène étudié. La seconde à construire un modèle théorique à partir de l'essence-même de la pratique étudiée, pour donner à voir et à comprendre une situation ou un phénomène spécifique. Dans ce cas, c'est bien la pratique évaluative qui contraint le modèle théorique. Enfin, la troisième posture engage une démarche dans laquelle prend forme une re-conceptualisation théorique du phénomène évaluatif à partir des données empiriques étudiées. Nous y revenons dans notre méthodologie.

⁴⁵ Les auteurs posent la même intention sur la modélisation, que nous traitons plus loin.

L'acte de produire un modèle à partir de données, ou d'en exploiter un existant pour les interpréter constitue donc un enjeu central dans la recherche de compréhension et d'interprétation des phénomènes évaluatifs. Comme le dit Figari (2012), il est fondamental de « disposer, dans le déroulement de l'activité évaluative, d'outils conceptuels [...] aptes à expliquer le résultat de différentes opérations éparses grâce à un comparant unique et reconnaissable » (p. 219). Dans le cadre de notre recherche, c'est le modèle de l'alignement curriculaire élargi que nous avons choisi pour interpréter les phénomènes évaluatifs révélés par nos données. Issu de la recherche sur les curricula, il conceptualise et modélise la cohérence recherchée entre les processus d'enseignement et d'apprentissage et le référentiel exploité (le curriculum), entre le référentiel et les démarches d'évaluation (ici, sommatives), et entre les processus d'enseignement et d'apprentissage et les démarches d'évaluation. C'est le fait de nous focaliser sur cette dernière dimension qui nous a contraint à le conceptualiser de manière élargie. Comme nous le précisons dans notre revue de littérature, la recherche dans ce champ, à notre connaissance, n'a que trop superficiellement abordé certaines étapes de la pratique évaluative sommative en jeu dans ce modèle que sont la pondération et la notation. Nous n'avons pas non plus trouvé assez de travaux significatifs étudiant en profondeur les processus évaluatifs saisis dans ce modèle.

Ce sont également ces constats qui nous ont poussé à penser que l'alignement curriculaire dans sa perspective élargie pourrait permettre une modélisation des phénomènes évaluatifs. Comme le disent Mottier Lopez & Figari (2012) : « par rapport au modèle, préconstruit par nature, la modélisation [...] constitue une procédure active, appliquée à la préparation d'une évaluation précise et locale, en relation avec une problématique qui, elle, est générale » (p. 16). De plus, elle représente de manière abstraite les relations existant entre les éléments constitutifs de l'évaluation et les facteurs qui les influencent (Figari, 1998). Dans ce sens, la modélisation poursuit un but compréhensif plutôt qu'explicatif, puisqu'elle participe à construire un cadre d'interprétation situé théoriquement sans lequel l'évaluation est en proie à une trop importante subjectivité et à l'arbitraire de l'évaluateur (Machado, 2007).

Dès lors, nous distinguerons ici le modèle de la modélisation en insistant sur le fait que le premier renvoie à un souci d'intelligibilité des phénomènes d'un point de vue théorique, alors que le second tente d'explicitier les relations que nous établissons entre les différentes unités conceptuelles du point de vue de la pratique évaluative réelle. Dans ce sens, nous présenterons un modèle de l'alignement curriculaire élargi pour expliciter et comprendre des phénomènes

évaluatifs qui ont effectivement lieu dans une perspective dynamique, et non pour normer la pratique en référence à des manières de faire considérées comme « bonnes » ou « justes ». En effet, c'est notamment au travers des questions de la pratique que nous avons étoffé, voire affiné théoriquement ce modèle. Nous sommes ici, comme le montre Alkin (2003), dans un modèle descriptif - et non prescriptif - dont le but est de fournir des éléments permettant de décrire, comprendre ou expliquer la pratique évaluative, ici, sommative. De ce fait, et suivant l'auteur, nous faisons l'hypothèse qu'un tel modèle offrirait la possibilité de « développer une théorie empirique » du concept d'alignement curriculaire élargi (p. 87).

Ceci étant, le recours à des modèles théoriques n'est pas exempt d'enjeux, d'un point de vue de recherche et de formation. Scriven (1991), dans une étude sur les liens entre la théorie de l'évaluation et la pratique évaluative, a affirmé : « theory is something of a luxury to an evaluator » (p. 360). Il a observé que la théorie, au travers, notamment, de l'exploitation de modèles à visées compréhensives, n'était que très rarement mobilisée par les évaluateurs. Des résultats que confirme Christie (2003) : « most evaluators do not use frameworks aligned with a specific theoretical model » (p. 33).

Ces éléments étant précisés, il nous incombe maintenant d'explicitier nos choix et nos ancrages théoriques. Nous commencerons par préciser ce que nous entendons par évaluation sommative.

2 Evaluation sommative ou certificative ?

Deux termes recouvrent les pratiques évaluatives qui dressent les bilans d'apprentissages réalisés par les élèves au terme d'une période donnée en vue d'aboutir à une certification le plus souvent chiffrée : l'évaluation sommative et l'évaluation certificative.

Bien que nombre de chercheurs francophones aient opté pour le second⁴⁶, nous avons opté pour le premier, non pas pour nous démarquer, mais pour des questions de cohérence que nous explicitons ici:

- L'évaluation sommative se définit de la manière suivante: «Elle a pour objet un inventaire. Son usage social est de vérifier (probation), sa fonction sociale de certifier. Elle a comme fonctions annexes classer, situer ou encore informer » (Hadji, 1989, cité

⁴⁶ C'est le cas notamment des membres de l'équipe EReD de l'Université de Genève, dirigée par Lucie Mottier Lopez, dont les travaux font référence aujourd'hui au sein de la communauté francophone des chercheurs en évaluation.

par Mottier Lopez, 2015, p.23). Cette définition met en exergue la question de la preuve de l'apprentissage, ou du moins questionne les conditions nécessaires pour que des preuves d'apprentissage puissent hypothétiquement exister dans une épreuve sommative. Cet élément est à nos yeux intéressant, car il rejoint la question du « quoi évaluer » et celle des modalités permettant cette évaluation. Par ailleurs, elle rend explicite la fonction sociale de certification, puisque ce type d'évaluation, de fait, donne un message à de multiples acteurs, bien au-delà de l'élève. Un intérêt supplémentaire à suivre cette définition est de considérer que l'évaluation sommative peut, à certaines conditions, poursuivre d'autres fonctions, comme classer ou informer, deux visées que nous serons amené à conceptualiser car fort en lien avec les pratiques évaluatives observées.

- La quasi-totalité des chercheurs dont nous nous inspirons et qui ont infléchi nos choix conceptuels sont anglo-saxons et parlent de *summative assessment*. Ils la définissent essentiellement au regard de sa temporalité et de sa fonction principale qui est de certifier. Ainsi, par exemple, Brookhart (2017) dit : « Summative assessment is assessment that is conducted after the learning has taken place to certify what has been learned » (p. 8).

Nous sommes conscient que l'évaluation certificative recouvre toute ou partie de ces caractéristiques. Nous savons également que le terme sommatif, étymologiquement, peut engendrer des malentendus, dans la mesure où nous pourrions entendre ici que toute épreuve visant à certifier serait un cumul de preuves, alors qu'il n'en est rien. En effet, les auteurs qui ont influencé notre travail peuvent parler d'évaluation sommative, tout en la conceptualisant au regard des apprentissages et des objets de savoir, et en se distançant des modèles de mesure quantitativistes.

3 Un modèle théorique phare : l'alignement curriculaire

Dans un premier temps, nous présenterons notre modèle de manière globale, puis en détail, en nous arrêtant sur chacune de ses composantes et sur les relations qu'elles entretiennent. Parallèlement, et comme soulevé dans notre revue de littérature, nous expliciterons la notion de cohérence, parce que l'alignement curriculaire élargi repose sur sa conceptualisation. Notre modèle abordant trois champs de recherche complémentaires – les curricula, l'évaluation et la notation – nous en expliciterons en continu ses originalités, en spécifiant notamment ce qui le

différencie d'autres modèles théoriques. Suivra une partie théorique sur le type de validité que nous sollicitons dans notre recherche. Traiter de validité est apparu comme fondamental dans notre démarche, d'une part parce que cohérence et validité se différencient théoriquement, et d'autre part parce que cette question de validité est omniprésente dans les références dont nous nous inspirons.

3.1 Une première conceptualisation de l'alignement curriculaire incomplète

En reprenant Cohen (1987, 1995) et Guskey (2003), Gauthier, Mellouki, Bissonnette et Richard (2005) définissent l'alignement curriculaire comme l'élément « visant à assurer une correspondance élevée entre le curriculum, l'enseignement et l'évaluation » (p. 24). Plus loin, ils précisent que le modèle ne concerne pas uniquement les contenus d'apprentissage, mais aussi sur « les opérations intellectuelles et les différentes catégories de connaissances qui y sont reliées » (p. 25). Par exemple, l'apprentissage porte-t-il sur de la restitution, de la synthèse ou de la création ? Les catégories de connaissances sont-elles de l'ordre des faits, des concepts, des stratégies, des démarches métacognitives ?

Malgré ces premiers éléments théoriques fondateurs, ce qui interpelle, c'est la manière avec laquelle le modèle d'alignement curriculaire est ici appréhendé, théorisé. L'acception qu'en ont ces auteurs est selon nous applicationniste, voire prescriptive. En effet, ils décrivent trois étapes pour sa mise en œuvre.

- La première consisterait à définir ce qui doit être maîtrisé par les élèves, en termes de performances (l'« *output* » : le quoi mesurer). Pour eux, il s'agit ici de répondre à la question : « comment peut-on observer et juger le degré de maîtrise de telle connaissance ou compétence en termes de comportements observables ou de manifestations ? » (Gauthier et al., 2005, p. 24). La *formulation des objectifs d'apprentissage et des critères d'évaluation* est ici un geste clé de cette étape, dont les finalités sont de clarifier le curriculum, de le concrétiser afin de faciliter son enseignement et son évaluation.
- La deuxième étape consisterait à *planifier les séquences d'enseignement et d'apprentissage* en rapport avec les objectifs du curriculum retenus de manière à amener tous les élèves à apprendre et progresser. La question ici est : « comment permettre aux élèves d'apprendre ce qui doit être maîtrisé ? ».

- Enfin, la troisième étape viserait à *évaluer les apprentissages effectués par les élèves*. Elle porte sur l'enseignement dispensé, sur ce « qui fait objet d'apprentissage en salle de classe » (p. 25). Selon eux, il est important que les tâches évaluatives soient reconnaissables et congruentes avec ce qui a été enseigné et avec le curriculum.

Gauthier et ses collègues schématisent ces principes de la manière suivante :

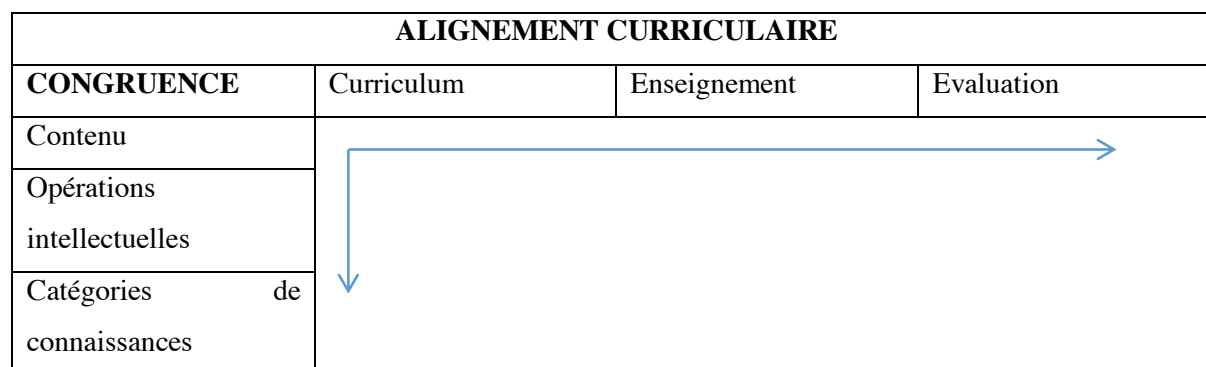


Figure 2 : Principe de l'alignement curriculaire selon Gauthier et al. (2005, p. 26)

Cette acception de l'alignement curriculaire interroge à trois niveaux :

- Le premier est relatif à la linéarité sous-jacente du modèle qui, de notre point de vue, *ne rend pas compte de la complexité de la pratique évaluative*. Là, c'est comme s'il suffisait de se fier étroitement à un enseignement inscrit dans un curriculum pour que les tâches évaluatives, de fait, soient alignées d'un point de vue curriculaire. Raisonner de la sorte revient à penser que tout processus d'enseignement suit une logique standardisée où les tâches d'apprentissage ressemblent aux tâches évaluatives. C'est loin d'être le cas, comme le montrent certaines recherches dans le champ (e. g., Bateman et al., 2009)⁴⁷.
- La deuxième interrogation tient au *manque de conceptualisation* qui entoure l'idée de congruence. Si les auteurs précisent qu'il est important de considérer les opérations intellectuelles comme faisant partie de l'alignement curriculaire, ils ne font jamais de lien explicite entre la congruence de ce dernier et le rôle que ces opérations intellectuelles y jouent, ni n'abordent les spécificités disciplinaires dans de telles démarches. Ils ne font pas non plus de lien explicite entre congruence et cohérence, alors que c'est cette dernière notion qui est théorisée dans une importante littérature dont nous faisons état dans la partie précédente.

⁴⁷ Pour prendre connaissance des résultats évoqués, consulter la revue de littérature en partie 2.

- Enfin, la troisième interrogation concerne *l'incomplétude théorique qui entoure l'évaluation*. Le raisonnement s'arrête au niveau des tâches évaluatives. Il n'est nulle part question du traitement de ces tâches, des informations sur les apprentissages des élèves qu'elles délivrent, ni de l'interprétation de ces dernières. Dans une perspective sommative, cela signifie que les processus de pondération et de notation ne sont pas théorisés, alors que, comme nous l'avons vu, ils sont indissociables de l'élaboration de l'évaluation en référence à un curriculum.

Notons toutefois que Gauthier et ses collègues ne sont pas les seuls à survoler conceptuellement cette question. Ils rejoignent selon nous Anderson (2002) qui, de notre point de vue, se situe dans les mêmes incomplétudes théoriques au niveau de l'évaluation sommative comprise dans ce modèle. En effet, il définit l'alignement curriculaire comme : « the level of coherence between curriculum objectives and assessment, between curriculum objectives and instructional activities, and finally between assessment and instructional activities » (p. 257). La validité du contenu, la façon dont il est enseigné, ainsi que les opportunités d'apprentissage de ce contenu offertes aux élèves au travers de tâches d'apprentissage font ici également partie du concept d'alignement curriculaire. L'auteur le schématise de la façon suivante :

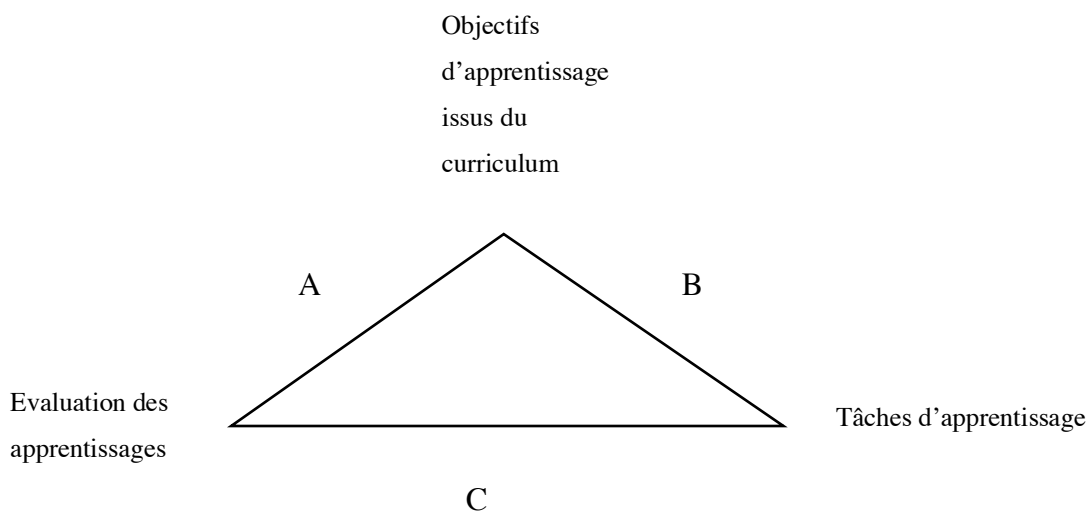


Figure 3 : L'alignement curriculaire selon Anderson (2002, p. 256)

Les questions formulées afin d'appréhender l'évaluation au sein de ce modèle sont ici aussi théoriquement incomplètes à nos yeux. Par exemple, au niveau de la relation A qui explicite le lien de cohérence entre l'évaluation des apprentissages et les objectifs d'apprentissage du

curriculum, Anderson propose la question correspondante suivante : « to what extent does the test measure the important curricular objectives ? » (p. 255), sans développer ce que sous-tend cette idée de mesure, ni la question des objectifs clés. Quant à la relation C, qui est censée mettre en évidence la cohérence qui devrait exister entre l'évaluation et les tâches d'apprentissage, l'auteur formule la question y relative : « is what we are teaching being tested ? » (p. 255). De nouveau, il reste très circonspect sur les questions centrales de pondération et de notation. Notons enfin que d'autres auteurs suivent ce développement du modèle avec les mêmes incomplétudes théoriques, en insistant surtout sur les caractéristiques du curriculum et des tâches, et de leur rapport de cohérence (e. g., Hammerness, 2006 ; Pellegrino, 2006 ; Martone & Sireci, 2009 ; Gagné, Dumont, Brunet & Boucher, 2013).

3.2 Une conceptualisation plus approfondie

Il faut se tourner vers Biggs pour voir un chercheur donner de l'importance à une forme de conceptualisation de l'évaluation dans le modèle de l'alignement curriculaire. Auteur-clé sur ce modèle, il insiste sur plusieurs dimensions de l'évaluation inscrite dans une recherche de cohérence curriculaire. Comme nous l'avons dit, les conclusions de ses travaux, bien que menés dans l'enseignement supérieur, sont tout à fait transposables dans l'enseignement secondaire et primaire. Ce qu'ils mettent en évidence est tout à fait intéressant d'un point de vue théorique. Biggs (1999) conceptualise ce qu'il appelle « *a constructive alignment* », qu'il définit de la sorte :

It is a fully criterion-referenced system, where the objectives define what we should be teaching; how we should be teaching it; and how we could know how well students have learned it. In aligned teaching, there is maximum consistency throughout the system. The curriculum is stated in the form of clear objectives which state the level of understanding required rather than simply listing the topics to be covered. The teaching methods chosen are those that are likely to realise those objectives; you get students to do the things that the objectives nominate. Finally, the assessment tasks address the objectives, so that you can test to see if the students have learned what the objectives state they should be learning. All components in the system address the same agenda and support each other. The students are "entrapped" in this web of consistency, optimising the likelihood that they will engage the appropriate learning activities. (p. 64)

Nous reviendrons plus loin sur cette définition, notamment lorsque nous aborderons la question des tâches évaluatives et des objectifs du curriculum que l'auteur explicite plus finement d'un point de vue théorique. Mais nous observons déjà que Biggs parle de l'alignement curriculaire

comme d'un système dont un des pôles est l'évaluation. Concernant les relations que l'évaluation entretient avec les objectifs du curriculum d'une part, et les tâches d'apprentissage d'autre part, Biggs et Tang (2009) insistent sur au moins trois plans :

- La perspective d'une évaluation qualitative ;
- L'utilisation de critères ;
- Une notation qui rend compte de l'apprentissage.

Au niveau de la pondération, il propose d'utiliser des « *grading criteria* », ou « *rubrics* » qu'il présente comme « the key to easy and successful grading » (p. 6). Qualitatifs, étroitement en rapport avec les contenus et objectifs, les critères rendent compte d'une perspective holistique de noter (« *holistic grading* ») que d'autres auteurs ont également développée (e.g., Royce Sadler, 2009) et qui vise à assurer la cohérence avec les tâches évaluatives et le curriculum. Nous y reviendrons. Plus loin, il développe sa conception de la notation, qui doit être en rapport avec l'apprentissage via les critères et les tâches, et qui permet de ce fait d'éviter les phénomènes de compensation⁴⁸. Dans ce sens, il montre combien la conception de l'évaluation prise dans le modèle de l'alignement curriculaire est incompatible avec la mesure (1999). Nous approfondirons également cet apport ultérieurement.

3.3 Quelques incomplétudes théoriques relatives à l'évaluation sommative

À la lumière de ces éléments, nous pourrions penser que Biggs a résolu la question de la cohérence de l'évaluation sommative au cœur du modèle de l'alignement curriculaire. Cela ne nous semble que partiellement correspondre à ses travaux. En effet, trois « angles morts » théoriques persistent selon nous, qui justifient la conceptualisation que nous présentons ici. Nous pensons notamment :

- À la question des disciplines et à leurs spécificités, surtout lorsqu'il s'agit de critérier certains objets de savoir particuliers en vue de noter des performances d'élèves ;
- Aux problèmes que soulève le passage des critères à la note lorsque celle-ci doit rester étroitement liée avec l'apprentissage évalué ;

⁴⁸ La compensation est présente lorsqu'un élève obtenant des points dans certaines tâches évaluatives et pas dans d'autres voit son résultat final être l'addition du tout. Il y a compensation car il peut rattraper des lacunes dans des dimensions de l'apprentissage évalué avec des acquis réalisés dans d'autres de ces dimensions. Le résultat prend alors la forme d'une moyenne qui ne permet plus de voir les réussites et les échecs.

- Au processus de construction de la note, notamment lorsque ce processus est confronté à des échelles amples⁴⁹ et imposées.

Autant d'éléments qui nous ont poussé à conceptualiser l'alignement curriculaire dans une perspective élargie à l'aide des apports de recherches menées sur la pondération et la notation.

4 Une nécessité théorique d'élargir le modèle : l'alignement curriculaire élargi

Il nous a donc été nécessaire de croiser certaines approches, afin de théoriser le système - « *the network* » - que représente l'alignement curriculaire dans une perspective élargie (Biggs, 2003). Comme le disent Trigwell et Prosser (2014) : « Evidence of the alignment between qualitatively different approaches to teaching and aspects of assessment can be found as a part of several studies, but to our knowledge has been the principal focus of none » (p. 145). Là, nous avons été surpris de constater que, souvent, les auteurs ne font qu'évoquer les rapports que l'évaluation sommative devrait entretenir avec les apprentissages réalisés ou avec le curriculum, mais qu'ils peinent à les théoriser.

En croisant ces apports, nous avons pu dégager deux caractéristiques du modèle de départ sur lesquelles nous nous appuyerons :

- L'alignement curriculaire rend compte de relations complexes existant entre le curriculum prescrit, les tâches d'apprentissage, les démarches d'évaluation, les processus de pondération et les pratiques de notation. Il doit dès lors être davantage considéré comme un système à plusieurs composantes qui s'influencent mutuellement qu'à un modèle mécaniste dont les étapes de résolution suivraient un ordre prédéterminé;
- Les processus d'enseignement et d'apprentissage sont au centre de la conceptualisation du modèle. Dans ce sens, on peut faire l'hypothèse que chacune de ses composantes et les relations qu'elles entretiennent entre elles doivent pouvoir en rendre compte. Ainsi, par exemple, on pensera l'élaboration de critères ou de tâches évaluatives en se référant aux objets de savoir qui ont été enseignés, aux situations d'apprentissages offertes aux élèves, aux objectifs réellement travaillés.

Dès lors, d'un point de vue théorique, conceptualiser une cohérence d'alignement curriculaire

⁴⁹ Nous entendons ici des échelles à multiples échelons.

(Anderson, 2002), ou encore un « *constructive alignment* » (Biggs, 1999, 2003, 2010) renvoie à un raisonnement tout sauf linéaire. C'est la raison pour laquelle nous proposons de conceptualiser l'alignement curriculaire (selon les auteurs cités) dans une perspective élargie (en référence à la pondération et à la notation) de la manière suivante :

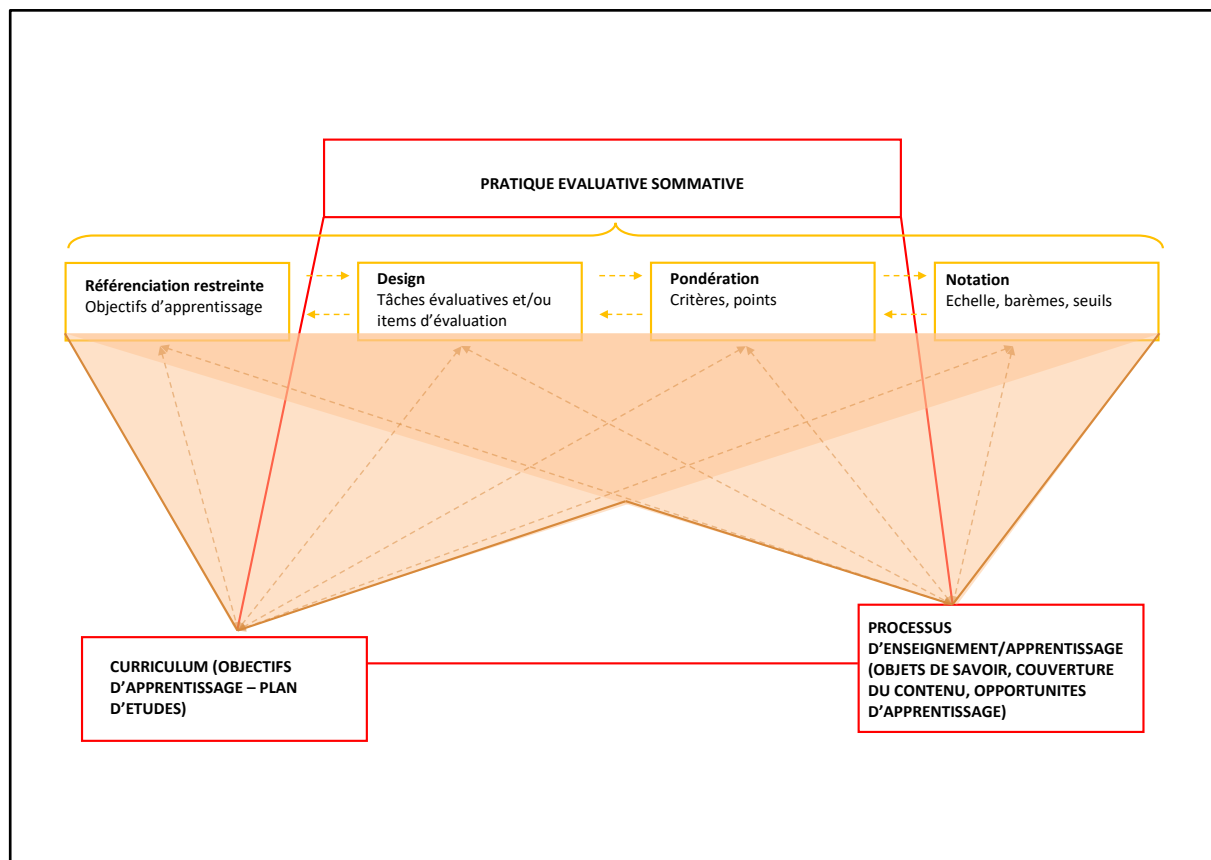


Figure 4 : L'alignement curriculaire élargi inspiré d'Anderson (2002) (Pasquini, 2016)

L'alignement curriculaire élargi prend ici selon nous un double statut. Il est un modèle qui offre une théorisation originale des phénomènes évaluatifs saisis dans une perspective sommative. Mais, soumis aux pratiques réelles et effectives des enseignants, il peut jouer un rôle de modélisation qui donne à voir les rapports de cohérence qui devraient exister entre chacune de ses composantes, et l'influence des processus d'enseignement et d'apprentissage sur ces dernières. Dans ce sens, il permet aux praticiens une réflexion située sur leurs pratiques. Nous détaillons maintenant chacune de ses composantes.

4.1 Un modèle devenant avant tout systémique

Notre modèle se lit en « entrant » par la triangulation mettant en relation le curriculum, les processus d'enseignement et d'apprentissage et la pratique évaluative sommative. Nous

retrouvons là la manière avec laquelle Anderson (2002) met en évidence les liens de cohérence qui doivent exister entre ces trois composantes. Il faut ensuite voir les composantes « Curriculum » et « Processus d'enseignement et d'apprentissage » comme ayant une grande influence sur la démarche d'évaluation sommative, déclinée de manière horizontale. Conformément au modèle de départ de l'alignement curriculaire, il est important que les objectifs d'apprentissage (référenciation restreinte⁵⁰), les tâches ou items d'évaluation (*design*), les critères et/ou les points (pondération) et enfin les échelles, barèmes et autres seuils (notation) soient en cohérence, d'où les flèches à double sens qui les relient. Mais cette horizontalité n'est qu'apparente, car elle s'inscrit dans le système plus général en faisant écho aux composantes « Curriculum » et « Processus d'enseignement et d'apprentissage ». Par exemple, si l'on considère ce modèle comme un système, nous observons que la pondération peut non seulement rendre compte de la valeur donnée aux apprentissages dans les tâches évaluatives et permettre une notation qualitative référée à l'apprentissage, mais qu'elle se réalise en référence aux objectifs et aux processus d'enseignement et d'apprentissages, omniprésents. Il en va de même pour les objectifs, qui, idéalement, s'inscrivent dans un curriculum qui en déterminera la valeur en référence aux processus d'enseignement et d'apprentissage. Nous parlons ici de perspectives « *bottom-up* » ou « *top-down* » d'alignement curriculaire élargi (Biggs, 1999 ; Walvoord & Johnson Anderson, 2009), qui demandent de penser la cohérence soit en partant du curriculum pour « descendre » vers les tâches évaluatives et ainsi de suite, soit en raisonnant depuis la notation, la pondération ou les tâches évaluatives, en « remontant » vers le curriculum.

D'un point de vue théorique, nous proposons donc un déploiement conceptuel de la pratique évaluative au sein même du modèle d'Anderson (2002). Cette nouvelle acception met en évidence que toute recherche de cohérence d'alignement curriculaire est tributaire des processus d'enseignement et d'apprentissage et du curriculum en vigueur, et que la pratique évaluative s'y inscrit jusque dans l'acte de noter, dans une perspective qualitative. Il est donc important maintenant de définir théoriquement ce que l'on entend par ces termes, fondamentaux. Ensuite, nous conceptualiserons chacune des composantes du modèle ainsi construit, et les liens qu'elles entretiennent avec leurs composantes périphériques, voire au-delà.

⁵⁰ Les concepts entre parenthèses qui caractérisent les composantes de l'alignement curriculaire élargi sont introduits ici. Par souci de concision et de clarté, nous avons choisi de les définir et de les développer plus loin.

4.2 Un modèle qui met en évidence la qualité de l'apprentissage dans l'évaluation sommative

L'enjeu de penser une évaluation sommative référée à l'apprentissage jusque dans la notation est central dans notre modèle, dans le prolongement des travaux de Biggs (2003). Nous montrerons que cet enjeu oriente conceptuellement nos choix à tous les niveaux. Aussi, nous proposerons de concevoir les objectifs issus du curriculum, les tâches évaluatives, les systèmes de pondération, les processus de notation et les relations de cohérence que toutes ces composantes doivent entretenir, comme un système dynamique devant tendre vers une mise en évidence de la qualité des apprentissages, et non de leur quantité.

En effet, quantifier l'apprentissage revient à le considérer comme un objet physiquement appréhendable, saisissable dans l'espace/temps, et à penser qu'une unité de mesure objectivement définissable peut lui être applicable afin de le saisir sous une ou plusieurs dimensions, chacune pouvant être pourvue d'une échelle numérique (Hadji, 2012). Ces caractéristiques, qui relèvent du paradigme de la mesure (De Ketele, 2012), ou du modèle de la mesure (Brookhart, 2005), ne sont pas compatibles avec l'apprentissage à nos yeux, et encore moins avec son évaluation⁵¹. Les croyances que ces conceptions véhiculent chez les enseignants sont en porte-à-faux avec une conception de l'apprentissage davantage orientée autour d'un processus dynamique, contextualisé et continu. Comment orienter les pratiques d'évaluation sommative vers une logique de progression des apprentissages, quand l'enseignant, comme le dit Biggs (2003), pense que la connaissance est une somme d'unités, que chaque unité en vaut une autre, qu'elles peuvent s'additionner, se compenser, que peu importe ce qui est correct pourvu qu'il y en ait assez, que ces unités sont les devises absolues des échelles, que l'évaluation ainsi élaborée est scientifique, précise, objective, ou encore que la distribution des acquis détermine les résultats ? À l'instar de Cardinet (1989), nous pensons donc que les possibilités de mesure de l'apprentissage en évaluation scolaire sont radicalement insuffisantes au regard des besoins des enseignants de documenter finement les parcours de leurs élèves. Dans ce sens, il importe de concevoir l'évaluation comme devant être prioritairement qualitative - c'est-à-dire en lien direct avec l'apprentissage, du point de vue des habiletés cognitives en jeu et des objets de savoir sur lesquels il porte - y compris quand il s'agit

⁵¹ Précisons que nous parlons d'évaluation sommative interne à la classe. Les épreuves externes, en revanche, peuvent s'inscrire dans le paradigme de la mesure.

d'attribuer des notes⁵².

Conceptuellement, la question de la qualité de l'évaluation renvoie à plusieurs paradigmes, dont celui de l'évaluation de la maîtrise des apprentissages (De Ketele, 2012), ou au modèle de l'évaluation (Brookhart, 2005), qui considèrent que tout apprentissage rend compte d'un processus cognitif complexe non standardisable. Être attentif à la qualité de l'apprentissage, c'est donner une importance cruciale aux informations que l'on récolte chez les élèves en vue de prendre des décisions relatives à leur progression, en référence à des objets de savoir disciplinaires, à des habiletés cognitives et aux caractéristiques du contexte. Et c'est au travers d'un jugement dont la référence est le modèle de l'alignement curriculaire que ce processus décisionnel prend forme ici. L'enjeu est de situer tout apprenant en référence à des objectifs, et de le mener le plus loin possible dans son cheminement cognitif. En termes d'évaluation sommative, cela signifie notamment de penser les procédures, les outils, les démarches évaluatives comme devant servir des diagnostics qualitatifs. Comme le dit Brookhart (2005) :

To assess the quality of your students' work, you need to know what assessment options are available to you, how to construct or select an appropriate assessment from these options, how to get these assessments to yield good-quality information, how to interpret the information and help students to interpret it, and how to use the information yourself and help students (and sometimes others) to use it. You also need to follow this cycle through to the end so that the information does get used; otherwise, the students' time and yours are wasted. (p. 7)

Le modèle de l'alignement curriculaire élargi s'inscrit donc selon nous pleinement dans cette visée qualitative. Il s'agit maintenant de voir, conceptuellement, comment nos choix théoriques ont poursuivi cette perspective.

5 Le curriculum, une référence parmi bien d'autres

Toutes les sources consultées citent le curriculum comme étant la référence à partir de laquelle les objectifs sont tirés ou re-formulés pour piloter l'enseignement, l'apprentissage et l'évaluation (e. g., Biggs, 2003 ; Bateman et al., 2009). Si ce postulat peut être envisagé de manière idéale, il ressort que du point de vue des pratiques et de la recherche, il n'en est rien. En matière d'évaluation, des travaux ont montré que bien d'autres références que le curriculum sont en jeu lorsque les enseignants traitent de la question des objectifs, comme les moyens

⁵² Ces éléments ont déjà été exposés dans notre revue de littérature, mais il nous semble ici important de les répéter, car ils constituent une des clés de voûte de nos choix conceptuels.

d'enseignement ou les documents internes à l'établissement (Mottier Lopez, 2016). C'est dès lors de multiréférentialité dont il s'agit lorsque l'enseignant formule des jugements ou prend des décisions (Mottier Lopez, 2008, 2015), notamment quand il énonce les objectifs qu'il pense évaluer. Toutefois, cette notion de curriculum reste centrale si l'on envisage que les pratiques d'évaluation sommative doivent idéalement s'y référer.

Nous l'avons montré dans notre revue de littérature, les définitions et acceptions du curriculum sont multiples et variées. Pour rester en phase avec notre cadre théorique tout en ne nous éloignant pas de notre contexte spécifique, nous nous baserons sur celles issues du monde anglophone qui, comme le souligne Lenoir (2011), « englobent le texte officiel (le curriculum prescrit), délibérément organisé et aménagé par l'enseignant (sa planification), ainsi que son actualisation en classe (le curriculum réel) » (p. 16). Ainsi, nous définirons le curriculum comme une « séquence d'opportunités d'apprentissages offertes aux étudiants afin de promouvoir l'apprentissage d'un contenu spécifique » (Schmidt, Mc Knight, Houang, Wang, Wiley, Cogan & Wolfe, 2001 cité par Crahay, 2006, p. 9). Cette définition nous permettra de nous centrer sur les questions d'évaluation sommative des apprentissages en référence à des contenus et des objectifs, quelles que soient leurs origines. C'est par ailleurs là, à notre sens, que s'incarne la question de la qualité des apprentissages. En effet, se référer aux contenus, cela oblige à privilégier l'« apprendre » au détriment du « faire » qui lui, renvoie à l'activité pour elle-même (par exemple, faire des exercices).

Dans le contexte de notre recherche, c'est le Plan d'études romand (PER) qui est le curriculum de référence, formulé en termes de compétences (Marc & Wirthner, 2013). C'est sous cette forme que ses concepteurs ont choisi de présenter ce qui est aujourd'hui le référentiel d'enseignement de tous les enseignants de Suisse romande.

6 Les objectifs d'apprentissage, ou la question de la référenciation

Cela étant, le PER n'est pas un référentiel d'évaluation. Il ne donne aucun renseignement sur les modalités, les temps, le contenu, la forme ou les critères des évaluations, ni sur les exigences relatives aux différents cycles ou disciplines. Le PER n'est donc qu'« un curriculum qui décrit ce que les élèves doivent apprendre » (PER, 2010, p. 20).

Il devient dès lors fondamental de théoriser son emploi au cœur du modèle de l'alignement curriculaire élargi lorsqu'on pense son exploitation au sein de la pratique évaluative sommative.

Il peut sembler paradoxal de se référer à un curriculum comme s'il était un référentiel d'évaluation alors que celui-ci n'a pas été conçu dans ce sens. Pourtant, ce sont bien les objectifs du PER que les enseignants doivent évaluer d'un point de vue sommatif, comme le souligne le cadre légal en vigueur dans le canton de Vaud⁵³.

Dès lors, il constitue bel et bien, et dans des degrés d'exploitation variables, une référence parmi d'autres. Toutefois, il nous a semblé risqué d'assimiler l'acte de se référer au PER à un processus de référentialisation. Comme l'ont montré certains auteurs (e. g., Figari & Remaud, 2014 ; Mottier Lopez, 2015), ce processus renvoie à toutes les étapes des démarches évaluatives dans leur acception large et non pas seulement à la formulation ou la sélection des objectifs dans le curriculum à des fins d'évaluation sommative. Malgré cela, les enseignants, à ce niveau, se réfèrent bel et bien à quelque chose lorsqu'ils concrétisent les apprentissages qu'ils poursuivent et évaluent en fin de séquences ! C'est la raison pour laquelle, afin d'explicitier ce processus, nous nous sommes inspiré du concept de référenciation développé par Vial (2012).

6.1 Vers une référenciation restreinte

L'auteur montre que tout évaluateur est amené sans cesse à faire des choix, sans se renier (1999/2008). Son action renvoie à une multitude de références. Il est dès lors acteur de leur sélection, de leur articulation et de leur usage. C'est ce que Vial nomme la référenciation, qui recouvre le fait de s'appuyer sur la pensée d'autres acteurs, l'action d'emprunter une théorie, l'intention de privilégier un angle de prise de vue spécifique sur un objet pour pouvoir en parler, ou encore le fait de s'imposer un cadre de pensée (Vial, 1999, 2012). Ce concept nous semble très pertinent pour comprendre et décrire « ce qui fait foi » chez les enseignants quand ces derniers opèrent des choix au niveau des objectifs d'apprentissages à conduire et évaluer d'un point de vue sommatif. Cela étant, le concept recouvre une ampleur importante de références possibles, alors que nous ne nous centrons ici « que » sur le rapport aux objectifs que les enseignants considèrent, qu'ils soient prescrits, reformulés par les professionnels ou issus de multiples sources. C'est la raison pour laquelle nous parlerons de référenciation restreinte.

6.2 Des objectifs d'apprentissage orientés théoriquement

Nous ne pouvons aller de l'avant sans expliciter le cadre théorique choisi pour définir ce que

⁵³ Voir http://www.vd.ch/fileadmin/user_upload/organisation/dfj/dgeo/fichiers_pdf/CGE2015.pdf.

nous appelons ici objectifs d'apprentissages, ces derniers étant une composante phare de notre modèle.

Bien que formulés dans le PER en termes de compétences⁵⁴, nous considérerons les apprentissages à évaluer issus du curriculum comme des objectifs⁵⁵. À ce niveau, nous suivrons de nombreux auteurs ayant travaillé sur l'alignement curriculaire (e. g., Biggs, 2003 ; Hammerness, 2006 ; Squires, 2009) qui prônent l'usage de taxonomies pour décrire leur niveau de complexité. Comme nous le justifions dans notre revue de littérature, notre choix s'est porté sur la taxonomie d'Anderson et Krathwohl (2001). Elle propose de considérer tout apprentissage au travers d'objectifs appariant une habileté cognitive à un contenu⁵⁶. En effet, pour parler d'apprentissage, l'habileté cognitive seule ne suffit pas, et l'unique contenu non plus. Conceptuellement, cette taxonomie croise six niveaux d'habiletés cognitives montant en complexité (se souvenir, comprendre, appliquer, analyser, évaluer et créer) avec quatre contenus de nature différente (les faits, les concepts, les procédures et les démarches métacognitives). Les objectifs se formulent en associant les deux. Comme le soulignent Mottier Lopez et Laveault (2008), au travers de cette taxonomie, « les auteurs ont pour ambition de démontrer les interactions entre enseignement et évaluation » (p. 9), ce qui est cohérent avec notre modèle. Cette perspective cognitive nous semble donc bien convenir à l'étude de la cohérence d'évaluations sommatives dans une perspective d'alignement curriculaire élargi. Tout d'abord, car elle a été éprouvée dans plusieurs recherches partant des pratiques réelles, comme la nôtre (e. g., Anderson, 2002 ; Bateman et al., 2009). Ensuite, car elle donne une importance cruciale aux caractéristiques disciplinaires des objets étudiés. Comme le précise Anderson (2002), citant Gomoran, Porter, Smithson et White (1997) : "Clearly, to predict student achievement gains from knowledge of the content of instruction, a micro-level description of content that looks at cognitive demands by [type of knowledge] is the most useful approach considered to date" (p. 258). Suivant ces derniers auteurs, nous faisons l'hypothèse que formuler ainsi des objectifs participe à souligner les aspects qualitatifs de l'apprentissage. Cette taxonomie offre donc une grille de lecture théorique de la cohérence de l'alignement curriculaire élargi, puisque c'est au travers de la complexité des habiletés cognitives appariées aux contenus qu'il est possible d'observer dans quelle mesure cette complexité de

⁵⁴ Le lecteur se référera à la revue de littérature pour l'argumentation de notre choix initial de nous distancer de cette approche.

⁵⁵ C'est d'ailleurs ce dernier terme qui est utilisé dans le PER.

⁵⁶ Pour rester fidèle aux auteurs, nous entendons le contenu comme tout élément faisant référence aux objets de savoir disciplinaires.

l'apprentissage perdure, des objectifs jusqu'au processus de notation.

Au cœur du déploiement conceptuel de la pratique évaluative prise dans l'alignement curriculaire élargi, il est dès lors important que les apprentissages et les objectifs qui s'y réfèrent soient évalués au travers de tâches significatives, en rapport avec les contenus et niveaux de complexité formulés. Conjointement, les tâches évaluatives doivent évaluer ce qui est formulé. Ce double mouvement est important à conceptualiser, car des travaux ont montré que c'est la nature de l'objectif d'apprentissage qui devrait déterminer la nature et le contenu des tâches évaluatives (e. g., Brookhart, 2011 ; McMillan & Nash, 2000). Ce qui nous amène à préciser ce que sont pour nous des tâches évaluatives.

7 Les tâches évaluatives sommatives : une question de *design*

Une littérature de référence pour nous sur l'alignement curriculaire traite de *design* pour parler du processus de construction des tâches évaluatives - *designing Assessment Tasks* : ATs (e. g., Biggs, 2003 ; Biggs & Tang, 2009)⁵⁷ - ou parfois de « *task-design processes* » (Klenowski & Wyatt-Smith, 2014, p. 50). Ce processus accorde une importance particulière à la cohérence qui devrait exister entre les objectifs d'apprentissage à évaluer et les tâches évaluatives.

Deux étapes y sont décrites comme importantes. La première, présentée comme la plus simple, consiste à sélectionner une tâche qui correspond à l'habileté cognitive exprimée au travers de l'objectif et au contenu sur lequel porte ce dernier. Un consensus émerge sur le fait que cette tâche doit être de complexité élevée, de manière à rendre compte de la qualité des apprentissages, et conjointement, permettre un traitement qualitatif des traces des élèves. Et c'est précisément ce traitement qualitatif qui constitue la deuxième étape de sa réalisation. Il est intéressant de noter que dans cette perspective, le *design* des tâches évaluatives comprend également la manière avec laquelle elles pourront être pondérées, puis notées. Nous traiterons ces deux points plus loin.

Dès lors, pour ces derniers auteurs, les tâches évaluatives assurent la cohérence d'alignement curriculaire quand elles rendent visibles les activités mentales - « *cognitive engagement* » (Biggs, 1999, p. 67) - que les élèves sont amenés à produire en référence aux habiletés des objectifs évalués. Biggs et Tang (2011) détaillent à ce niveau l'importance de produire des

⁵⁷ Il ne faut pas confondre cette acception avec celle d'autres chercheurs, comme Wiggins et McTighe (2005), qui parlent du *design* comme le processus de construction de toute la démarche d'apprentissage.

tâches évaluatives également en rapport avec les contenus abordés et qui respectent la taxonomie : par exemple, si l'apprentissage s'est déroulé à des niveaux d'analyse de sources scientifiques, il n'est pas pertinent d'évaluer d'un point de vue sommatif leur description ou identification.

7.1 Des tâches authentiques

Ces auteurs parlent de tâches authentiques (*authentic tasks*). Mottier Lopez (2015), dans un article où elle problématise la différence entre tâches d'apprentissage et d'évaluation, montre que cette question de tâches authentiques n'est pas récente. Reprenant les travaux de Wiggins (1988), elle explicite les différences entre les évaluations traditionnelles (*typical tests*) et les tâches authentiques (*authentic tasks*). Nous reprenons ces éléments ici, car ils concordent en grande partie avec notre conceptualisation des tâches.

Tableau 1 : Differences between typical tests and authentic tasks (Wiggins, 1988) cité par Mottier Lopez (2015, p. 77)

Typical test	Authentic tasks
Require correct responses only	Require quality product and/or performance, and justification
Must be unknown in advance to ensure validity	Are known as much as possible in advance ; involve excelling at predictable demanding and core tasks ; are not « gotcha ! » experiences
Are disconnected from a realistic context and realistic constraints	Require real-world use of knowledge : the student must « do » history, science, etc. in realistic simulations or actual use
Contain isolated items requiring use or recognition of known answers or skills	Are integrated challenges in which knowledge and judgment must be innovatively used to fashion a quality product or performance
Are simplified so as to be easy to score reliably	Involve complex and non-arbitrary tasks, criteria, and standards
Are one shot	Are iterative ; contain recurring essential tasks, genres and standards
Depend on highly technical correlations	Provide direct evidence, involving tasks that have been validated against core adult roles and discipline-based challenges
Provide a score	Provide usable, diagnostic (sometimes concurrent) feedback : the student is able to confirm results and self-adjust as needed

Nous observons ici que les tâches authentiques devraient notamment :

- Permettre une mise en évidence de la qualité de l'apprentissage ;
- Se centrer sur les objets disciplinaires ;
- Viser le développement de la connaissance ;
- Intégrer des défis d'apprentissage ;
- Evaluer des niveaux d'apprentissage de complexité élevée ;
- Etre constituées de tâches emblématiques de la discipline ;
- Offrir des possibilités de feedback.

Nous sommes conscient que la question de l'authenticité des tâches peut prêter didactiquement à controverse, dans le sens où l'on peut interroger la pertinence d'inscrire chacune d'entre elles dans des situations de la vie quotidienne⁵⁸, des « *key challenging situations* » (Wiggins & McTighe, 2005). Pour éviter de nous disperser, nous n'engagerons pas ici ce débat. Selon nous, cette caractéristique des tâches authentiques n'enlève toutefois en rien l'intérêt de la conceptualisation proposée par ces chercheurs, puisque nous y retrouvons certaines caractéristiques du modèle de l'alignement curriculaire élargi : centration sur les contenus, approche qualitative de l'apprentissage et de l'évaluation, cohérence avec les critères d'évaluation, cohérence taxonomique. Par ailleurs, d'autres éléments ouvrent des perspectives théoriques intéressantes : la question des tâches emblématiques des disciplines, qui renvoie au statut qu'elles donnent aux différents objets de savoir disciplinaires qui y circulent (par exemple, la production écrite en français, ou la résolution de problèmes en mathématiques) ; et la question des feedbacks inscrits dans une perspective sommative, qui met notamment en lumière la problématique de la nature du retour fait à l'élève (Harlen, 2005).

7.2 Quelles caractéristiques théoriques des tâches évaluatives ?

Dans cette perspective, deux constats peuvent être opérés. Tout d'abord, nous considérerons la tâche évaluative comme étant possiblement constituée de sous-tâches, mais amenant toutes le plus possible à des réponses ouvertes demandant un développement de la part de l'élève (Royce Sadler, 2009). Cela exclut l'acception d'une tâche évaluative sous forme de succession d'items. En effet, l'item renvoie à une question de basse complexité et à un traitement de l'évaluation dans le modèle de la mesure (Anderson, 2002 ; Biggs, 2003) ou à des évaluations standardisées

⁵⁸ Ceci d'autant plus que la forme scolaire, dans son essence même, se distancie des situations de la vie quotidienne en didactisant toute tâche d'apprentissage ou d'évaluation.

(Pellegrino, 2006 ; Martone & Sireci, 2009). Ensuite, à l'instar de notre modèle théorique, nous pensons que notre conceptualisation de la tâche marque le lien entre tâche d'apprentissage et tâche évaluative, puisque ce sont les mêmes préoccupations théoriques qui les définissent. En cela, nous abondons dans le sens de Mottier Lopez (2015) qui affirme : « l'apprentissage de l'élève ne peut s'appréhender – et donc s'évaluer – que dans la saisie de l'activité de l'élève en interaction avec la tâche, elle-même saisie dans des situations et réseaux de significations plus larges » (pp. 62-63). C'est ici bel et bien l'alignement curriculaire élargi qui guide et oriente le sens que prend la tâche évaluative.

En sus, et comme nous l'avons déjà évoqué, une des caractéristiques fondamentales de la tâche évaluative au cœur de notre modèle est qu'elle peut être traitée de manière qualitative. En d'autres termes, elle doit donner à voir des dimensions de l'apprentissage évalué, de manière à mettre ce dernier en évidence en lui assignant des systèmes de valeur eux aussi qualitatifs. Cela nous amène à traiter maintenant de la pondération.

8 La pondération ou la question de la valeur de l'apprentissage

En évaluation sommative, nous définirons la pondération comme le fait de signifier l'importance d'un apprentissage - son poids - au travers d'outils, comme les critères ou les points (Marzano, 2002). Considérée comme une pratique fondamentale, la pondération, comme nous l'avons mis en évidence dans notre revue de littérature, a été, et est encore de manière paradoxale, très peu étudiée dans la recherche en évaluation sommative. Peut-être parce que, comme le suggèrent Mottier Lopez, Tessaro, Dechamboux et Morales Villabona (2012) « il s'agit d'un aspect de la pratique évaluative qui reste peu accessible à la conscience des enseignants ou qu'ils ne désirent pas forcément rendre publique » (p. 10) ? Toujours est-il que sa conceptualisation mérite développement.

Il existe deux modalités de pondération de l'apprentissage qui s'appuient sur des outils distincts : l'attribution de points et la formulation de critères. La première, quantitative, renvoie à une évaluation le plus souvent par items, qui permettent de comptabiliser les points jusqu'à des totaux qui servent ensuite à formuler des notes calculées à l'aide de pourcentages ou de sommes de points. Biggs (2003) traduit cette manière de faire de la façon suivante : « a process that converts nonquantitative datas into numbers » (p. 5). En affirmant cela, il dénonce, encore une fois, l'emprise du modèle de la mesure sur les pratiques et montre le paradoxe de vouloir quantifier l'apprentissage, ce qui n'a, malgré son succès dans la pratique, guère de sens, comme

nous l'avons précédemment montré. En revanche, la seconde modalité de pondération à l'aide de critères, qualitative, s'inscrit pleinement dans le modèle de l'alignement curriculaire élargi, comme nous l'avons déjà évoqué. Etant donné que ce modèle, et *a fortiori*, son acception élargie, poursuivent une recherche de mise en valeur de la qualité de l'apprentissage, il va de soi que la pondération ne peut se réaliser qu'au travers de critères, seuls outils à même de relever le défi. Là, deux perspectives de recherche se distinguent. Celle qui considère qu'un appariement entre critères et points est possible (Royce Sadler, 2009). Et une autre qui exclut les points du raisonnement porté sur la pondération et l'évaluation (Biggs, 2003). Là, le chercheur précise qu'il s'agit de « making holistic judgement that match outcomes with qualitative criteria » (p. 5). C'est cette dernière modalité que nous conceptualiserons.

8.1 Les critères d'évaluation : l'unique outil pour mettre en valeur la qualité de l'apprentissage

Pour élaborer une conceptualisation des critères cohérente avec notre modèle, nous avons choisi de nous focaliser sur les travaux qui ont étudié leur inscription dans une cohérence d'alignement curriculaire. Notre tâche n'a toutefois pas été aisée. En effet, comme le précisent Reddy et Andrade (2010) : « important aspects of validity have not yet been addressed at all, including the need to establish the alignment between the criteria on the rubric and the content or subject being assessed » (p. 445). Les chercheurs travaillant spécifiquement sur l'alignement curriculaire problématisent peu cette question des critères au sein du modèle. Peut-être est-ce dû au fait qu'il semble évident de considérer que tout critère devrait être aligné curriculairement avec les tâches évaluatives, ou qu'objectifs et tâches sont indissociables pour définir les critères (Jonsson, 2014) ? Et pourtant. La recherche peine à investiguer cette question (Panadero & Jonsson, 2013). Ces manques nous ont donc demandé une analyse de différentes approches dans le domaine, afin de choisir une conceptualisation des critères en phase avec notre cadre théorique. Dans cette perspective, notre choix s'est porté sur la définition de quatre catégories de critères d'évaluation pour problématiser la question de la cohérence au sein de notre modèle théorique, à des fins de mise en valeur de la qualité de l'apprentissage évalué (Moskal & Leydens, 2000).

8.1.1 Les critères centrés sur les tâches (*task-specific rubrics*)

Ces critères ont la particularité d'entretenir avec les tâches évaluatives un rapport très étroit, parfois même de terme à terme avec les consignes. Ils ne peuvent donc pas être généralisés à

des familles de situations évaluatives et demandent des ajustements, voire des reconstructions permanentes, car ils mentionnent des éléments issus des tâches souvent très descriptifs et fort précis - par exemple, « présence d'un titre en rapport avec la thématique du roman » en production écrite de résumé - (Brookhart, 2005). De ce fait, ils sont facilement utilisables, permettent une attribution de points rapide. En revanche, ils sont tellement explicites qu'ils ne peuvent être communiqués aux élèves, sous peine de bachotage, ou d'amplification du phénomène de « *teaching to the test* » (Popham, 2001). Dans cette perspective, ils ont tendance à perturber une cohérence d'alignement curriculaire élargi, car conceptuellement, ils sont proches d'une conception quantitative de la pondération puisqu'ils appellent les points et renvoient à un niveau de complexité bas dans l'apprentissage. Nous les considérerons toutefois ici car ils sont fréquemment utilisés par les enseignants et donnent à voir ainsi certaines contraintes liées à la pratique.

8.1.2 Les critères analytiques (*analytic rubrics*)

Très étudiés, ces critères ont été conceptualisés par divers auteurs. Pour Jonsson et Svingby (2007), dans une pondération par critères analytiques, « the rater assigns a score to each of the dimension being assessed in the task » (p. 132). La tâche est donc disséquée, et l'on donne à chacune de ses parties un critère auquel on attribue ensuite un nombre de points - par exemple, toujours en production écrite de résumé « cohérence globale de l'introduction ». Toutefois, nous observons que malgré la dimension qualitative de la démarche, l'évaluateur perd de vue la globalité de la tâche, et donc de l'apprentissage qui y est évalué, les critères analytiques ayant la caractéristique d'être considérés séparément les uns des autres (Brookhart, 2005). Royce Sadler (2009) met en lumière les mêmes phénomènes. Pour lui, le fait de produire un jugement qualitatif chaque fois spécifique en fonction du critère est typique d'une pondération de type analytique, il parle alors de « *criterion-by-criterion judgments* » (p. 3). Meier, Rich et Cady (2006) complètent en précisant que ces critères représentent tous les aspects de la solution attendue pour être en phase avec les points attribués à chacun d'entre eux. À ce niveau, c'est un moyen pertinent pour identifier les forces et faiblesses des élèves.

Mais d'autres travaux, tout en soulignant l'efficacité de ces critères, montrent qu'ils comportent des limites lorsque les différentes dimensions qu'ils évaluent sont mises en commun pour établir un résultat final (Waltman, Kahn & Koency, 1998). Nous verrons que cet apport est fondamental dans notre recherche, puisqu'il s'agit ici de problématiser le lien de cohérence à construire entre la pondération et la notation.

Nous ne sommes donc pas ici face à un système de pondération qualitatif idéal, car les critères analytiques, dans leurs fondements, appellent plutôt une résolution finale de la performance sous forme de points. Toutefois, nous donnerons une importance à ces critères, car ils sont très fréquemment utilisés par les enseignants, mettent en évidence des dimensions qualitatives de l'apprentissage, et sont souvent considérés comme les plus efficaces (Royce Sadler, 2009). Cela étant, des conceptualisations de critères encore davantage orientées vers des dimensions qualitatives de l'apprentissage existent.

8.1.3 Les critères holistiques (*holistic rubrics*)

Ces critères se démarquent des précédents. Ils traduisent une autre perspective de l'évaluation sommative (Nitko, 2001), se caractérisent par le fait qu'ils amènent à une production de jugement plus global (Biggs & Tang, 2009 ; Isenhour & Kramlich, 2008 ; Klein et al., 1998). Là, les critères d'évaluation sont considérés simultanément, englobent plusieurs dimensions de la production attendue ou de l'apprentissage évalué. Ils ont l'avantage de fournir des repères plus généraux - par exemple, toujours en production écrite de résumé « cohérence des informations sélectionnées, de leur articulation et de leur synthèse ». Ils sont considérés unanimement comme plus adéquats que les critères analytiques pour une exploitation dans des évaluations sommatives, car ils permettent une notation plus aisée⁵⁹ et rendent mieux visible la complexité des apprentissages (Brookhart, 2005), ce qui demande toutefois aux enseignants de déterminer en amont les éléments qualitatifs attendus dans les réponses. C'est le mode de pondération que met également en avant Royce Sadler (2009), pour qui l'enjeu de produire un jugement global est premier. Selon lui, le seul moyen de rester en cohérence avec la complexité de l'apprentissage est de considérer le travail de l'élève comme un tout, « *an integrated entities* » (p. 3). La difficulté réside dans le fait que des critères holistiques requièrent une description verbale des caractéristiques essentielles des contenus évalués, démarche qui pose régulièrement problème aux enseignants (Meier et al., 2006). Ils n'ont donc pas que des avantages : leur niveau de généralité ne cible pas les informations en vue de donner des feedbacks aux élèves, et des biais importants peuvent apparaître dans leur exploitation, comme des effets de fatigue ou de halo dans les phases de corrections des épreuves (Royce Sadler, 2009).

Nous voyons donc qu'une conceptualisation de critères qualitatifs pour aller vers une cohérence

⁵⁹ Ce constat vaut toutefois surtout lorsque le nombre d'échelons est réduit.

d'alignement curriculaire élargi demande encore à être approfondie. C'est la raison pour laquelle nous devons tenter une conceptualisation complémentaire des critères, quelque peu hybride.

8.1.4 Les critères holistiques élargis (*annotated holistic rubrics*)

Se référant aux contraintes de la pratique évaluative en mathématiques et en anglais langue première au secondaire, Meier et al. (2006) entrevoient la possibilité de conjuguer les apports des critères analytiques et holistiques. Selon eux, ces critères (*annotated holistic rubrics*), que nous traduisons par critères holistiques élargis, prennent d'abord appui sur les caractéristiques des critères holistiques, tout en permettant, en fonction des apprentissages évalués, d'en spécifier les contenus :

A hybrid of the analytic and holistic rubrics, which helps teachers apply the same standards to all student papers, is an annotated holistic rubric. This type of rubric allows solutions to be scored holistically, while including general strengths and weaknesses that support the rating. In addition to the annotated holistic rubric or general scoring rubric, which is not task specific, specific scoring rubrics can be created that apply the annotated holistic rubric to specific constructed-response tasks. (p. 70)

Cette proposition est à nos yeux conceptuellement intéressante. Elle s'inscrit pleinement dans notre modèle, puisqu'elle articule une recherche de cohérence qualitative au regard des tâches évaluatives et des contenus évalués avec le raisonnement qui va présider à la construction de la note. En effet, cette conceptualisation des critères tente de tenir compte à la fois de la complexité de l'apprentissage évalué et des contenus spécifiques en jeu. Dans ce sens, il est délicat d'en fournir des exemples, puisque l'enjeu est de les penser à la fois au regard des tâches et des objets, et en fonction de ce qu'ils éclairent dans les travaux des élèves. D'autres auteurs, comme Royce Sadler (2009) soutiennent également cette approche, mettant en évidence que tout processus de critérisation se réfère d'abord aux contenus évalués, et que dès lors, les critères ne peuvent être entièrement analytiques ou holistiques. Par ailleurs, cette approche est fidèle aux deux caractéristiques que doit posséder une démarche critériée de ce type : mettre en évidence la qualité de l'apprentissage et offrir une stratégie de pondération (Reddy & Andrade, 2010). Elle offre aussi la possibilité de rendre compte des processus d'apprentissage en jeu et des contenus jugés importants, ce qui est fondamental dans notre modèle (Parke, 2001).

Cela étant, nous faisons ici l'hypothèse que ces derniers critères pourraient avoir un inconvénient majeur : leur rapport à des échelles de notation à multiples échelons. Nous y

revenons plus loin.

8.2 La question du poids donné aux critères

Le processus de pondération caractérise les actions qui déterminent le poids (*weight*) assigné, ici, aux différents critères. Pour Marzano (2002), c'est à ce niveau que la forte variation entre évaluateurs observée par certains travaux peut être réduite. Pour comprendre l'impact que cette question a sur la cohérence de l'alignement curriculaire élargi, il nous importe d'en détailler les fondements théoriques, sachant que deux manières d'attribuer du poids aux critères ont été conceptualisées à ce jour à notre connaissance.

8.2.1 Les « *rate-weight-sum* »

La première se nomme « *rate-weight-sum* » (Royce Sadler, 2009), que nous traduisons comme la somme des poids relatifs. Dans une telle démarche, l'évaluateur attribue à chaque critère un nombre de points sur une échelle correspondant à son importance, au regard des apprentissages. Il est ainsi amené à évaluer chaque critère en référence à son échelle. Une pondération supplémentaire (un coefficient) peut encore donner davantage de poids à l'un ou l'autre critère, que l'on multipliera par le nombre de points donnés. La somme des points est ensuite convertie en une note. Ce système correspond à une perspective analytique, avec les avantages et inconvénients déjà mis en évidence plus haut. Il serait, sans avoir recours aux points, plus adapté à des démarches d'évaluation formative, dans la mesure où des critères analytiques permettent mieux de formuler des feedbacks (Harlen, 2005). Par ailleurs, et comme déjà énoncé, l'idée de somme aboutissant à une seule note produit des effets de compensation entre critères qui diminuent la cohérence du verdict final (Biggs & Tang, 2009).

8.2.2 Les « *rubrics* »

La deuxième manière de faire se nomme « *rubrics* ». Nous parlerons ici de critères qualitatifs⁶⁰. Dans cette perspective, chaque critère est décrit de manière qualitative :

Each cell in a given row typically contains descriptive text that spells out the characteristics of a particular level or 'standard' for that criterion. This text can take several forms, including 'verbal quantifiers' and 'typical features'. The verbal-quantifier form can use a variety of words, but all amount to much the same

⁶⁰ Royce Sadler explique qu'il existe plusieurs dénominations à ce niveau. Nous utiliserons ici celle de critères qualitatifs « *qualities criteria* » afin d'éviter les malentendus.

thing. They range in meaning from the equivalent of ‘little if any evident’ to ‘an extremely high level present’. The typical-features form of cell entry consists of a qualitative description of the corresponding ‘standard’, often with reference to sub-attributes of the main criterion. Regardless of the form of the text in the cells, the assessor nominates the cell that best characterizes the quality of each student work on each criterion. (Royce Sadler, 2009, p. 5)

Deux manières de procéder pour établir la note sont ensuite possibles. Soit l’évaluateur fait la somme des points assignés à chaque critère puis transforme ce total en une note – ce qui correspondrait à ce que le chercheur nomme « *verbal-quantifier form* ». Soit il procède à un jugement global, en déterminant les seuils requis pour chaque critère, qu’il transforme ensuite en note. C’est à notre sens ce qui distingue cette approche critériée des autres, puisque chaque critère, dans ce dernier cas de figure, est évalué à l’aune d’un seuil qui intervient dans l’établissement de la note. Et c’est bien ce passage entre des critères holistiques et la construction de la note qui reste selon nous encore assez délicat. Dès lors, nous faisons l’hypothèse que c’est à ce niveau que la cohérence de l’alignement curriculaire élargi pourrait être mise à mal, notamment parce que l’approche induit un comptage de critères. Certes, ce dernier renvoie à des dimensions qualitatives de l’apprentissage, mais qui, à un moment donné, aboutissent à une notation obligatoirement objectivable au travers d’échelons. Une tension théorique est donc ici bien présente, qu’il s’agira de mettre à l’épreuve de nos données issues de la pratique. Il demeure néanmoins que cette théorisation est bien éloignée du modèle de la mesure, et permet d’envisager l’évaluation sommative dans une perspective qualitative, jusqu’à la notation, au cœur d’une cohérence d’alignement curriculaire élargi.

Il reste maintenant à conceptualiser cette problématique de la notation, étroitement liée à celle des critères et de la pondération en général (Brookhart, 2017).

9 La nécessité de conceptualiser différemment la notation

La conceptualisation de la notation que nous adoptons suit celle que nous avons choisie depuis le début. Nous cherchons à théoriser une notation à visée qualitative, référée à l’apprentissage. En d’autres termes, nous exposons ici une conceptualisation de la notation alignée curriculairement avec un système de pondération et un *design* de tâches évaluatives également qualitatifs, rendant compte d’apprentissages réalisés dans un système centré sur les processus d’enseignement et d’apprentissage et référée à un curriculum. À ces fins, nous parlerons de notation holistique. Ce concept intègre la notion d’échelle, nous la traiterons donc conjointement.

9.1 Noter et évaluer : quelle articulation ?

Pour rester fidèle à notre modèle, nous conceptualiserons la notation comme étroitement liée aux processus d'évaluation tout en s'en distinguant, les deux étant également saisis et intégrés au cœur des processus d'enseignement et d'apprentissage. En effet, leur articulation est au service de prises de décisions les plus cohérentes possible. Les premières peuvent prendre de multiples formes (diagnostique, formative, sommative), s'appuient sur des *design* de tâches évaluatives spécifiques et contextualisés au regard des réalités de classe. Les pratiques de notation, si elles doivent être en cohérence avec elles, sont en revanche en partie dépendantes de normes extérieures à la classe, de prescriptions, et relèvent de mécanismes souvent idiosyncrasiques, ce qui demande de les théoriser séparément, tout en les considérant comme faisant partie du système d'évaluation inscrit dans l'alignement curriculaire élargi (Biggs, 2003). McMillan et Nash (2000) ont modélisé le rapport étroit entre les pratiques d'évaluation et de notation, en mettant en évidence que si elles sont toutes deux tributaires de facteurs comme les valeurs des enseignants, les réalités de classe ou les éléments contextuels, elles se distinguent au niveau des prises de décision. Il y aurait donc, théoriquement, une distinction à opérer dans la nature des jugements que posent les enseignants quand ils évaluent d'un point de vue sommatif et quand ils notent.

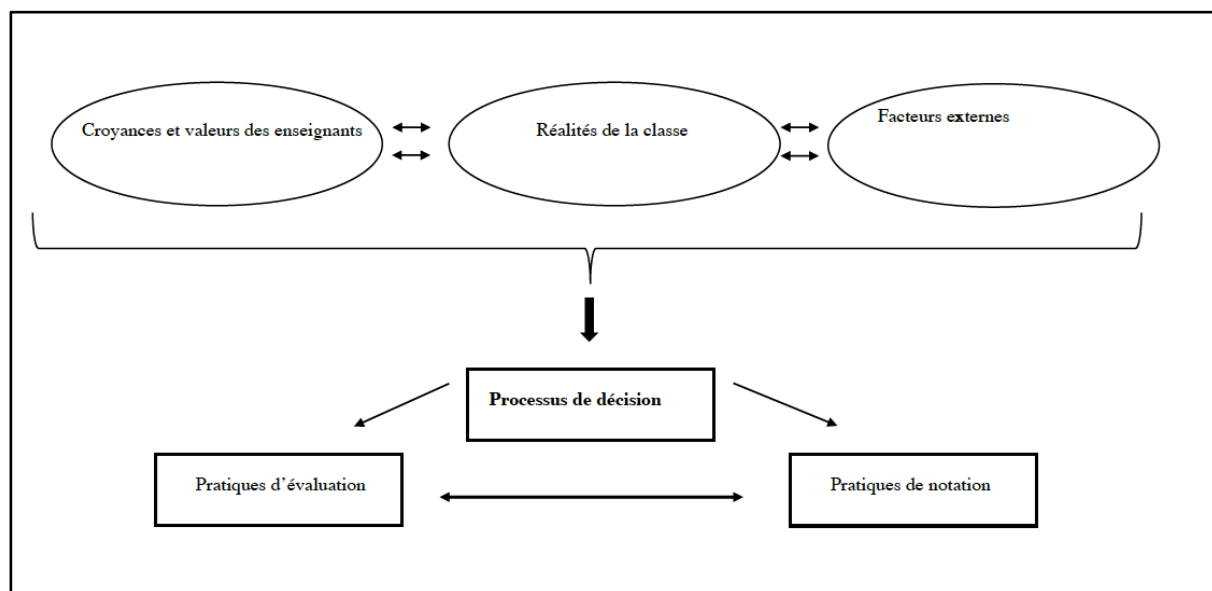


Figure 5: Le processus de décisions des enseignants au travers des pratiques d'évaluation et de notation selon McMillan et Nash (2000, p. 8)

Cet élément est particulièrement intéressant pour nourrir l'approche conceptuelle sous-tendue par notre modèle. Elle tendrait à considérer que d'un point de vue de pratique, un enseignant

peut développer des pratiques évaluatives très cohérentes dans l'alignement curriculaire jusqu'aux tâches évaluatives voire jusqu'à la pondération, mais que cela n'assure en rien la cohérence de ses pratiques de notation dans une perspective élargie. En d'autres termes, il est possible d'observer des pratiques « alignées » entre les objectifs et les tâches évaluatives, voire la pondération, puis pouvant être potentiellement « désalignées » jusqu'à la note (Airasian & Miranda, 2002)⁶¹. C'est à nos yeux un postulat que défendent nombre de chercheurs ayant beaucoup étudié les pratiques de notation, comme Brookhart (2017), lorsqu'elle affirme par exemple : « recent research suggests that we should not expect graded achievement and tested achievement to be the same thing » (p. 19). Il est dès lors important de caractériser cette phase de la notation plus finement.

9.2 La notation comme un processus

La notation peut être définie comme le « fait d'associer à un travail ou à une performance une note, c'est-à-dire une valeur numérique comprise entre des bornes conventionnelles » (Dubus, 2008, p. 269). Nous nous distancierons de cette définition, qui, à nos yeux, assimile conceptuellement la notation à un produit figé, résultant d'opérations mécaniques et arithmétiques. Pour rester fidèle à notre orientation conceptuelle, nous considérerons plutôt la notation comme « a complex context-dependant process that serve multiple roles » (Walvoord & Johnson Anderson, 2009, p. 2). Cette définition est conceptuellement fondamentale à nos yeux. Tout d'abord, car elle pose la notation comme relevant de la cohérence d'un système complexe, dont le modèle compréhensif serait ici l'alignement curriculaire élargi. Ensuite, car elle la définit comme un processus et non comme un produit fondé sur l'exploitation d'outils standardisés, le plus souvent externes à l'apprentissage qui a eu lieu en salle de classe. En d'autres termes, le processus de notation doit pouvoir rendre compte de l'apprentissage évalué, au même titre que n'importe quelle autre étape de la pratique évaluative, et est dès lors chaque fois spécifique à la situation. À ce titre, il devrait :

- Fournir des renseignements cohérents sur les apprentissages des élèves au travers de feedbacks ;
- Transmettre, comme vecteur de communication, des informations exploitables à tous les acteurs concernés (élèves, enseignants, parents, directions) ;

⁶¹ Ces auteurs parlent de « *misalignment* » (p. 254).

- Encourager la motivation des élèves, puisqu'il met en évidence leurs forces, mais aussi les apprentissages encore insuffisants à améliorer ;
- Aider à l'orientation et aux transitions en permettant une centration sur les apprentissages ;
- Encourager la régulation des apprentissages.

Au sein de notre modèle, la notation correspond dès lors à une transaction (Hadji, 2016) à multiples facettes, interdépendantes. Elle fait donc intimement partie de la cohérence de l'alignement curriculaire élargi. Raison pour laquelle nous parlerons de notation holistique.

9.3 Une notation, de fait, holistique

Biggs et Tang (2009) expliquent les raisons pour lesquelles seule une notation holistique peut s'inscrire dans une cohérence d'alignement curriculaire élargi : « curricular alignment requires a change from a quantitative and analytic mindset, that does atomise knowledge into marks, to one that uses and grades assessment tasks *qualitatively* and *holistically* wherever possible » (p. 2). Le principe fondamental ici est de donner à voir la qualité de l'apprentissage dans la notation. Le processus de notation doit dès lors être en cohérence avec un système de pondération s'appuyant sur des critères également holistiques. L'enjeu est notamment ici de faciliter la pratique évaluative de l'enseignant, et de donner davantage d'informations à l'élève. Conceptuellement, il s'agit donc de mieux décrire le processus de notation (Tierney, 2015), afin de mieux en comprendre les caractéristiques et sources d'altération, surtout quand ces dernières sont évoquées par les enseignants (Brookhart, 2011).

La notation holistique implique de considérer le travail de l'élève dans l'évaluation sommative comme relevant de l'intégration de divers éléments complexes (Royce Sadler, 2009). Dans ce sens, elle est le fruit d'un jugement global, s'appuyant sur des critères qui rendent compte de la complexité des contenus et des apprentissages. Certains chercheurs, comme Isenhour et Kramlich (2008), ont mené des travaux en mathématiques dans lesquels la note holistique était produite suite au jugement global de l'enseignant sur des critères pondérés au travers de trois positions (+, $\sqrt{}$, -). D'autres, à l'instar de Klein et al. (1998), ont mis en oeuvre des processus de notation prenant appui sur des échelles holistiques (*absolute holistic scale* : les points sont assignés aux critères sur une échelle allant de 0 à 100, où 65 détermine l'échec, 70 le seuil marginal, 75 le seuil de suffisance, etc.), ou encore des échelles holistiques relatives (*relative holistic scale* : par exemple, une échelle par critère de 1 à 5, où 1 signifie très en deça du seuil

de suffisance, et 5 très au-delà du seuil de suffisance). Mais ces travaux ont surtout analysé la fiabilité des processus de pondération en vue de construire une note holistique, et moins les enjeux de construction de la note elle-même saisie dans un tel processus. C'est peut-être une des raisons pour lesquelles, selon nous, ces approches ressemblent davantage à des processus de notation analytiques, puisque l'évaluation de l'apprentissage suit une perspective visant à quantifier chaque critère, puis à produire la note en référence à cette quantité (Brookhart, 2005).

9.3.1 Une notation holistique sans points, est-ce possible ?

D'un côté donc, la contrainte de « transformer » des critères, même holistiques élargis, en une seule et unique note est encore et toujours conceptuellement délicate. De l'autre, notre modèle d'alignement curriculaire élargi ne peut être cohérent qu'au travers d'une notation éminemment qualitative. Il s'agit donc pour nous de tendre vers une conceptualisation de la notation sans points, basée uniquement sur des critères qualitatifs. C'est ce que proposent nombre de chercheurs comme Biggs (1999, 2003), Biggs et Tang (2009, 2011) ou plus récemment Brookhart (2017), en exemplifiant plusieurs manières de procéder. L'une d'entre elles consiste à croiser les objectifs évalués avec les tâches évaluatives, de manière à repérer quels apprentissages sont évalués dans quelles tâches. La procédure la plus simple correspond à un appariement une tâche-un objectif. Mais la situation peut se complexifier lorsqu'un objectif est évalué au travers de plusieurs tâches.

Tableau 2: Le processus de notation holistique selon Biggs et Tang (2011, p. 7)

Objectifs	Tâche 1	Tâche 2	Tâche 3	Tâche 4
- relate	√	√		
- applique		√	√	
- identifie			√	
- explique				√
- communique				√
Attribution des notes				
A	1 et 2 « A », autres objectifs B			
B	Tous les objectifs B/C et un « A »			
C	Si 3, 4 et 5 atteints			
D	Si seulement 2 objectifs sont atteints			

Conceptuellement, cette avancée est pour nous significative, car elle met au centre la cohérence qui existe entre les objectifs évalués, les tâches évaluatives et le processus de notation, dans une

recherche de mise en évidence de la qualité des apprentissages. Elle met également en suspens la question de l'attribution des points, dont nous avons déjà soulevé les limites.

Toutefois, en suivant les propositions des auteurs cités, des interrogations théoriques demeurent. Nous citerons notamment l'action d'évaluer au sein de la même épreuve des objectifs de complexité variable, la tendance à valider chaque objectif dans une première phase avant d'arriver à la note finale, le fait de ne pas référer les objectifs à des contenus en vue de formuler des critères, ou, dès lors, l'aspect aléatoire qui peut se dégager au travers des caractéristiques des notes (Royce Sadler, 2009). Un autre aspect questionne et représente selon nous une limite importante de leur raisonnement : le nombre d'échelons à disposition pour valider l'épreuve. Il est donc important de rester vigilant.

9.3.2 Une notation à conceptualiser encore davantage

Ces limites relèvent trois incomplétudes que nous traiterons plus loin pour asseoir encore mieux nos choix théoriques.

- La première est relative à l'aménagement de telles démarches dans des contextes où des échelles amples sont prescrites. En effet, comment conceptualiser une notation holistique au travers d'un nombre important d'échelons devant chacun idéalement rendre compte des différentes dimensions de la qualité de l'apprentissage évalué ? A partir de combien d'échelons cette progression, cette description, deviennent-elles irréalisables ?
- La deuxième incomplétude concerne le rôle que jouent les caractéristiques des disciplines dans ces processus de notation, dans la mesure où elles impactent fortement la construction des critères référés aux contenus disciplinaires et aux objectifs. Y a-t-il des disciplines, voire des objets de savoir disciplinaires qui sont plus aisément notables que d'autres dans une perspective de notation holistique ?
- La troisième aborde la question du jugement professionnel mis en oeuvre dans l'élaboration des outils et des décisions évaluatives, et plus particulièrement des éléments qui participent à sa construction. Quels en sont les références ? Notre modèle nous oblige à théoriser ces trois points.

9.4 Des échelles holistiques⁶² ?

La fonction première de l'échelle, rappelons-le, est, comme le souligne DeLandsheere (1992), d'« ordonner des individus par rapport à une qualité qu'ils possèdent à un degré plus ou moins élevé » (p. 146). C'est donc bien la nature de cet ordonnancement qui est en jeu dans cette conceptualisation. Le processus de notation devenant qualitatif, l'échelle de notation traditionnelle, élaborée à l'aide d'algorithmes mathématiques chaque fois différents ou avec des outils externes à l'apprentissage calculant des seuils en référence au total de points prévus dans l'épreuve, n'a ici plus lieu d'être. L'évaluateur voit donc son questionnement sur la construction de la note se réorienter, en passant de « à quel pourcentage fixe-t-on le seuil de réussite de chaque note ? » à « quels apprentissages doivent être maîtrisés minimalement pour obtenir telle ou telle note ? ». Ce renversement est capital, mais fort délicat à opérer. En effet, un des rôles des échelles à visée holistiques est de fournir des feedbacks, en mettant en évidence pour l'élève et l'enseignant, outre la note, les apprentissages réussis et ceux qui méritent régulation. En cela, elles demandent aux enseignants des connaissances approfondies des objets disciplinaires (Mottier Lopez & Laveault, 2008) et des connaissances en évaluation, ce dont ils manquent, le plus souvent, cruellement (Moss, 2013).

Mais c'est notamment parce que la notation holistique s'appuie sur une nouvelle forme d'ordonnancement au cœur de l'échelle que certains auteurs en proposent dans lesquelles les notes sont directement critériées de manière progressive (e.g., Brookhart, 2005, 2011 ; Royce Sadler, 2009). Les critères sont identiques, globaux, et c'est leur contenu, allant du moins complexe au plus étoffé, qui détermine chaque note, de la plus basse à la meilleure. On retrouve ici la perspective holistique élargie dont nous avons parlé pour les critères, saisie au cœur des échelles. Cette proposition est fort pertinente, et tout irait bien si elle était transposable à toutes les disciplines de la scolarité. Mais cette problématique est bien plus complexe qu'elle en a l'air.

9.5 Le rôle des disciplines dans les processus de notation holistiques

Des différences dans les processus de notation en fonction des disciplines ont été mises en évidence dans certains travaux (e.g., Cross & Frary, 1999 ; Zhang & Burry-Stock, 2003). Un résultat qui ressort de ces études est directement en rapport avec notre problématique : les

⁶² Ce concept correspond à ce que certains chercheurs appellent des échelles descriptives (e. g., Laveault & Miles, 2008 ; Mottier Lopez & Laveault, 2008), ou des « *proficiency scales* » (Marzano & Heflebower, 2011).

pratiques de notation holistique sont plus délicates chez les enseignants de mathématiques que chez leurs collègues de langue, en partie à cause des caractéristiques des objets de savoir disciplinaires (Meier et al., 2006).

Ces constats, même s'ils ne peuvent être généralisés, sont éclairants. En effet, le modèle de l'alignement curriculaire élargi repose sur les caractéristiques des contenus évalués, qui, idéalement, participent à sa cohérence, de la formulation des objectifs jusqu'à la notation. Dans ce sens, nous pensons que conceptuellement, ces caractéristiques et leur appréhension par les enseignants jouent un rôle médiateur entre les pratiques de notation holistiques et l'expertise disciplinaire dont ces derniers peuvent rendre compte (Biberman & al., 2011). L'enjeu ici est de comprendre jusqu'où il est possible, au travers de notre modèle, de mettre en évidence les caractéristiques des contenus évalués dans les pratiques de notation. En d'autres termes, il nous semble important d'éprouver notre modèle à l'aune des disciplines, en vue, notamment, de définir les conditions disciplinaires, voire didactiques, d'une notation holistique. Ceci d'autant plus quand la discipline impacte aussi profondément sur les pratiques : « Deep knowledge of the domain of discipline and the expected outcomes in the intended curriculum are foundational to professional practice » (Klenowski & Wyatt-Smith, 2014, p. 57). Cette intention poursuit également une finalité de recherche à visée épistémologique, car il semble que des disciplines a priori proches comme les mathématiques et les sciences induisent, contre toute attente, des pratiques d'évaluation et de notation contrastées. Par exemple : « Science teachers seem to be more similar to their language-teaching colleagues and different from mathematic teachers in terms of both grading style and perception of the discipline » (Biberman-Shalev et al., 2011, p. 838).

Saisie au cœur d'une exigence d'expertise disciplinaire, dépendante du contexte d'évaluation et notamment des prescriptions en la matière, la question d'une notation holistique inscrite dans une cohérence d'alignement curriculaire élargi gagne encore en complexité et doit être éprouvée dans les pratiques réelles. Ceci d'autant plus que se pose la question du jugement global qu'induit cette approche : sur quels éléments repose-t-il ? Comment se construit-il ?

10 Vous avez dit jugement professionnel ?

Une foisonnante littérature montre que toute pratique évaluative s'appuie sur un jugement professionnel. Nous retiendrons qu'il est :

un processus qui mène à une prise de décision, laquelle prend en compte différentes considérations issues de l'expertise (expérience et formation) professionnelle de l'enseignant. Ce processus exige rigueur, cohérence et transparence. En ce sens, il suppose la collecte d'informations à l'aide de différents moyens, la justification du choix des moyens en lien avec les visées ou intentions et le partage des résultats de la démarche dans une perspective de régulation. (Lafortune & Allal, 2008, p. 27)

Par ailleurs, nous suivons Allal et Mottier Lopez (2007) qui en définissent quatre caractéristiques. Tout d'abord, il comprend deux démarches : réunir et confronter des sources diverses d'information ; les combiner à travers un raisonnement ou une interprétation (notamment quand il y a une construction de notes). Ensuite, il s'appuie sur des procédures d'évaluation que l'enseignant élabore et applique à l'ensemble des élèves, mais une de ses fonctions est de permettre des adaptations pour les situations singulières. Par ailleurs, il est un acte individuel, mais il s'élabore par des interactions et des communications dans un contexte social et institutionnel donné. Enfin, il implique une mise en relation entre les pratiques pédagogiques de l'enseignant, ses valeurs personnelles, la culture de l'établissement et les exigences du système scolaire (Allal & Mottier Lopez, 2008). On le voit, le jugement professionnel poursuit bel et bien une recherche de mise en évidence de la qualité de l'apprentissage. Mais que devient-il saisi dans notre cadre conceptuel ?

11 Et concevoir un jugement curriculaire holistique ?

Ce jugement professionnel, référé à l'alignement curriculaire élargi, devrait donc devenir à son tour holistique. A priori, cela ne pose pas de problème conceptuel. En effet, nous concevons tout à fait que la recherche de cohérence, ici aussi, s'élabore autour de sources d'informations pertinentes afin de prendre des décisions. Nous faisons également l'hypothèse que les conceptualisations récentes du jugement professionnel (voir l'introduction de l'ouvrage de Mottier Lopez et Tessaro, 2016) font référence, implicitement certes, à l'alignement curriculaire. Toutefois, dans une certaine mesure, nous pensons que notre acception du jugement en approfondit l'explicitation, notamment parce qu'elle s'ancre dans une référence systématique aux critères d'évaluation (Royce Sadler, 2009), eux-mêmes porteurs d'éléments de contenus et situés dans des processus d'enseignement et d'apprentissage inscrits dans un curriculum. Le jugement ainsi construit pourrait donc être caractérisé également de curriculaire, car il s'inscrit dans une recherche de cohérence entre diverses composantes de notre modèle et centrée sur les apprentissages et les contenus. Ainsi, lorsque l'enseignant récolte des informations, il agit de manière première au regard de la cohérence qui inter-relie les objectifs,

les habiletés cognitives, les objets de savoir, les tâches évaluatives, les systèmes de pondération ou encore les modalités de construction de la note. La dimension holistique de son jugement résiderait dans le fait qu'il est éminemment qualitatif et systémique, mais surtout lu au travers d'un modèle théorique, ce qui est de notre point de vue conceptuellement relativement nouveau. En effet, des travaux ont souligné les faiblesses de l'approche holistique, notamment au travers de jugements portés à la légère et éloignés des apprentissages (Bloxham, Boyd & Orr, 2011). Peut-être que cette acception pourrait y remédier, et ainsi apporter de nouveaux éléments pour conceptualiser la validité de l'évaluation ?

12 Une recherche permanente de validité

La question de la validité est régulièrement au centre des études sur les pratiques évaluatives sommatives et sur les processus de notation holistiques. Elle conditionne par ailleurs la compréhension des « bonnes » pratiques évaluatives sommatives inscrites dans une recherche de cohérence d'alignement curriculaire élargi. Dans ce sens, la validité est un des indicateurs principaux permettant de juger de la qualité des pratiques évaluatives sommatives (Brookhart, 2003). Nous la définissons ici comme étant *interne* à la situation choisie : « validity refers to how well scores relate to the teaching objectives and to the target performance domain » (Biggs, 2003, p. 9). Cette conceptualisation renvoie aux relations de cohérence qui devraient exister entre les objectifs, le système de pondération et les performances attendues. En d'autres termes, elle questionne la cohérence des informations récoltées sur les apprentissages des élèves dans les démarches d'évaluation sommative au regard des intentions de départ. Dans ce sens, nous retrouvons notre postulat affirmant qu'il est important d'être au clair sur les contenus évalués et sur le niveau de complexité des apprentissages en jeu (Brookhart, 2005).

L'acception de la validité inscrite dans notre modèle théorique est donc elle aussi systémique, car elle dépend de multiples facteurs qui interagissent entre eux. Comme le formulent Klenowski et Wyatt-Smith (2014) :

« There is now also widespread recognition that validity is not determined by a single statistic or test result, but rather by a body of information that demonstrates the relationship between the test and the behavior it is intended to measure. From this vantage point, the effectiveness and quality of assessment is vital » (p. 54).

Dans cette perspective, la validité rend compte d'un processus évaluatif et de notation résultant de décisions situées et non d'application d'outils de mesure. Aussi, nous parlons de la validité

de la tâche (son contenu évalue-t-il ce que l'évaluateur pense évaluer ?), de la validité de la pondération (les critères sont-ils formulés en référence au contenu et au niveau de complexité des objectifs évalués ?) et de la validité de la note (rend-elle compte des apprentissages réussis et de ceux qu'il est encore nécessaire d'approfondir au regard des processus d'enseignement et d'apprentissage ?).

On remarque donc que dans cette conceptualisation, la question de la validité de la notation est centrale. C'est par son truchement que les processus de notation peuvent être considérés comme cohérents, ou pas (Stiggins, 2001). Comme le précise Allen (2005) : « Validity is important because the sole purpose of grades is to accurately communicate to others the level of academic achievement that a student has obtained » (p. 218). Cette dimension met en lumière une volonté, chez certains chercheurs, de conceptualiser la validité en référence aux contraintes de la pratique également. En effet, partant du constat que toute situation d'évaluation sommative est spécifique à un contexte chaque fois différent, et qu'elle est soumise à des contraintes comme la notation, il importe de penser la validité en ces termes, éloignée d'un principe de standardisation. Moss (2003), à ce titre, affirme : « validity needs to be conceptualized, not just in the "context of classroom assessment" but in the contexts of the classrooms in which the assessment occurs » (pp. 20-21). Pour cela, elle va même jusqu'à proposer de penser la validité au travers de l'analyse de cas rendant compte de situations d'évaluation ayant donné satisfaction. Cette proposition empirique nous semble très pertinente. Nous pensons toutefois que la validité mériterait davantage d'être éprouvée à l'aune d'un modèle théorique, comme dans notre recherche.

Cette partie a exposé nos choix conceptuels relatifs à l'étude des pratiques évaluatives des enseignants et montre que notre modèle s'inscrit dans une épistémologie théorique de l'évaluation, puisqu'il s'ancre dans des cadres conceptuels spécifiques.

Il est toutefois encore nécessaire de présenter la perspective théorique que nous avons choisie pour étudier le développement des connaissances et compétences des enseignants en matière d'évaluation, à savoir l'*assessment literacy*. La prochaine partie s'y emploie.

Partie 4 - Cadre conceptuel sur l'*assessment literacy*

1 Introduction

Notre problématique porte sur les connaissances et compétences en évaluation en voie de développement chez les enseignants qui ont participé à notre recherche. Le but de cette problématique est triple :

- Mettre en évidence et thématiser les compétences évaluatives des enseignants qui hypothétiquement se développent par un travail interactif et réflexif sur le modèle de l'alignement curriculaire élargi ;
- Montrer dans quelle mesure les connaissances liées au modèle de l'alignement curriculaire élargi sont mobilisées dans les pratiques d'évaluation sommative réelles sous forme de compétences;
- Donner à voir à quels niveaux les pratiques de notation tendent à évoluer lorsqu'elles sont théorisées à l'aide du modèle de l'alignement curriculaire élargi.

Pour conceptualiser le développement des compétences évaluatives et l'acquisition des connaissances des enseignants dans ce champ au travers d'une perspective théorique adaptée, nous avons choisi *l'assessment literacy* (AL), qui a pour grand intérêt de cibler cette problématique.

Les travaux en AL se développent intensément dans la recherche en évaluation depuis les années 1990 (DeLuca, LaPointe-McEwan & Luhanga, 2015). Son but est de mettre en évidence les connaissances et compétences nécessaires à tout enseignant pour devenir un évaluateur cohérent, et de fournir des outils permettant d'observer le développement de ces compétences⁶³. Nous commencerons donc à en esquisser les contours. Puis nous nous positionnerons pour préciser ce que l'on entendra par *literacy*. Dans un deuxième temps, nous présenterons les idées directrices de l'AL et leur évolution dans divers courants de recherche, ce qui aboutira à en

⁶³ Précisons d'emblée que nous ne reprendrons pas ici le débat vif existant en éducation autour des compétences dont la définition peine à être stabilisée (Crahay, 2006 ; Legendre, 2008). Toutefois, par souci de rigueur, nous suivrons la définition de la compétence de Le Boterf (2001) qui dit : « un savoir-agir fondé sur la mobilisation et l'utilisation efficace d'un ensemble de ressources, tant internes qu'externes ». Elle nous semble pleinement en accord avec la conception qu'en ont les chercheurs en AL qui ont théorisé la question des ressources. Concernant les connaissances, nous pensons important de questionner lesquelles sont indispensables à l'enseignant (Paquay, 2012). Dans ce sens, nous postulons, en suivant toujours Paquay, Wouters et Van Nieuwenhoven (2010), qu'elles sont les unités d'apprentissage mobilisées par les acteurs sous forme de compétences (p. 10). Dans ce sens, nous considérerons que l'articulation entre connaissances et compétences relève moins d'une accumulation de ressources (ici, de connaissances) que de leur subtile mobilisation en vue de résoudre des situations complexes (Paquay et al., 2010). Nous développerons ce point dans le texte.

présenter les postulats et caractéristiques principales. Ce travail nous amènera à distinguer deux acceptions aux épistémologies contrastées : une AL orientée autour de la mesure des compétences développées, que nous qualifierons de *statique*, et une autre, qui sera la nôtre, considérant cette perspective théorique comme un outil pour décrire l'évolution des pratiques, dès lors plus *dynamique*. Cette dernière explicitation mettra en évidence l'influence de la dimension contextuelle sur toute pratique évaluative sommative ainsi que les mouvements à l'œuvre chez les enseignants lorsqu'ils sont invités à s'approprier un modèle théorique pour penser leurs pratiques évaluatives sommatives jusqu'à la notation. Nous montrerons ensuite comment nous avons abouti à la formulation de compétences en évaluation pour notre recherche, au nombre de six. Nous terminerons cette partie conceptuelle en esquisant les potentielles avancées que l'exploitation de cette perspective théorique pourrait nous faire réaliser dans l'interprétation de nos données.

2 L'*assessment literacy* : une perspective théorique pour comprendre et instrumenter le développement des compétences en évaluation

Depuis les années 90, une ample littérature anglo-saxonne s'intéresse aux connaissances que les enseignants devraient maîtriser et aux compétences qu'ils seraient censés posséder pour comprendre les phénomènes évaluatifs auxquels ils sont confrontés et ainsi développer des démarches d'évaluation valides et soutenant l'apprentissage des élèves. Ces connaissances et compétences constituent le cœur conceptuel de l'*assessment literacy*.

Dans un numéro thématique de la revue *Assessment in Education : Principles, Policy & Practice* traitant spécifiquement des compétences évaluatives des enseignants, DeLuca et Johnson (2017), en introduction, réaffirment ce que Popham disait déjà en 2009 : « *Assessment literacy is seen as a sine qua non for today's competent educator* » (p. 121). En s'appuyant sur plusieurs textes issus de diverses communautés de chercheurs des quatre coins du globe, ils souhaitent remettre en débat le fait qu'évaluer ne peut se réaliser sans la maîtrise de connaissances et au travers de compétences clés. Ils appellent à problématiser ce que l'on entend par compétences et connaissances en évaluation, et à réfléchir aux conditions pour envisager l'AL non seulement comme un outil de mesure des compétences, mais davantage comme une perspective théorique permettant d'accompagner le développement des pratiques évaluatives dans une épistémologie dynamique.

2.1 Comment comprendre *literacy* ?

La traduction conforme de *literacy* est « littératie », définie originellement comme « l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités » (OCDE, 2000, p. X d'introduction). Nous observons que cette définition générique fait référence à la nécessité, pour tout individu :

1. De maîtriser des connaissances (en langue, dans la définition de l'OCDE) ;
2. De les utiliser en situation ;
3. D'exploiter ses acquis pour atteindre certains buts, notamment en fonction des contextes dans lesquels ils sont poursuivis ;
4. De mettre à profit ses acquis pour le développement de connaissances et de compétences futures.

De ce point de vue, la littératie traite donc des connaissances et compétences que la personne possède relativement à un objet, afin de le comprendre et de l'exploiter à des fins personnelles. C'est peut-être la raison pour laquelle nous assistons au développement du concept dans divers domaines et champs : littératie numérique (pour comprendre, utiliser et créer des outils et médias numériques), littératie en santé (pour comprendre et utiliser des informations médicales afin de prendre des décisions de traitements) ou encore, et de manière non exhaustive, littératie visuelle (pour comprendre et utiliser tous les symboles autres que l'écriture).

2.2 Une littératie en évaluation ?

Hypothétiquement, il existerait donc des connaissances à posséder pour être un évaluateur compétent, cohérent, et donc capable de poursuivre des finalités liées à sa mission et prenant en compte son contexte. C'est en tous cas le postulat que défendent de nombreux chercheurs. Par exemple, Popham (2009) met en évidence une question clé que la formation à l'AL doit traiter : « What kinds of assessments do teachers most need to understand ? » (p. 5). Il montre que cette capacité à comprendre ce qui se joue en évaluation ne peut se réaliser sans une forme d'expertise qui repose sur l'acquisition de connaissances mises à l'épreuve de la pratique. Volante et Fazio (2007) vont dans le même sens. Pour eux, l'« *assessment literate teacher* » (p. 752) est capable de reconnaître et d'utiliser différentes formes d'évaluation (diagnostique, formative, sommative), considérées comme des stratégies d'enseignement.

Le but des démarches de formation inscrites dans la perspective de l'*assessment literacy* est donc de fournir aux enseignants des connaissances en évaluation pouvant être mises en œuvre dans la pratique afin de développer des compétences (Lukin, Bandalos, Eckhout & Mickelson, 2004). Si cette finalité est largement partagée par les sources consultées, l'épistémologie dans laquelle s'inscrivent les différentes conceptualisations de l'*assessment literacy*, en revanche, divergent.

2.2.1 Un premier lot de définitions de l'*assessment literacy*

Avant d'entrer dans les finesses conceptuelles de l'AL, il nous semble important de montrer en quoi certaines de ses définitions en traduisent des orientations épistémologiques et conceptuelles contrastées. Notons toutefois d'emblée que, paradoxalement, la définition générale de l'AL fait relativement consensus. Nous ne sommes pas ici confrontés, comme le dit Popham (2011), à une marée de définitions divergentes.

Cette définition première de l'AL, qui a inspiré bon nombre de chercheurs, est celle de Stiggins (1995, 2002) : « the capacity to know the difference between sound and unsound assessment » (p. 240). Elle rejoint les enjeux que la perspective théorique met en évidence : la possibilité, pour tout enseignant, de discriminer une évaluation soutenant l'apprentissage d'une autre plus arbitraire, moins cohérente. Plusieurs chercheurs s'ancrent clairement dans cette première acception (e. g., Mertler, 2004, 2009 ; Popham, 2004 ; Volante & Fazio, 2007). Pour Mertler (2004), citant Paterno (2001), elle est « the possession of knowledge about the basic principles of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment...and familiarity with alternatives to traditional measurements of learning » (p. 50-51). Pour Gotch et French (2014), reprenant Popham (2011), elle constitue « an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions » (p. 14). Pour ce dernier enfin, l'AL s'apparente à « someone's knowledge about educational tests and their roles » (p. 267). Ces définitions, de notre point de vue, valorisent l'acquisition de connaissances, mais thématisent peu leur exploitation dans des situations pratiques, même si certaines d'entre elles évoquent l'importance pour l'enseignant de prendre des bonnes décisions. Mais sur quoi ces décisions se basent-elles ? Quelle articulation existe-t-il entre la maîtrise des connaissances et leur exploitation dans la pratique pour une évaluation au service des apprentissages des élèves ? Comment observer l'acquisition des connaissances et le développement des compétences chez les enseignants ?

Autant de questions qui nous font penser que ces premières définitions sont incomplètes. En effet, nous faisons ici l'hypothèse qu'elles considèrent l'AL comme une perspective théorique relativement prescriptive, qui désignerait les connaissances et compétences que tout enseignant devrait posséder pour être un « bon » évaluateur. Pour le savoir, il suffirait donc de pré et post-tester ses connaissances et compétences avant et après une formation ad hoc pour mesurer ses progrès, et ainsi montrer qu'il devient *assessment literate*. C'est ce que propose notamment Mertler (2009)⁶⁴.

2.2.2 Une première épistémologie de l'*assessment literacy* plutôt statique

La plupart des chercheurs dont nous parlons ici considèrent que les connaissances et compétences en évaluation se réfèrent à des standards à atteindre et peuvent se mesurer. Nous faisons l'hypothèse que leur définition de l'AL s'en ressent, se réduisant à des listes de connaissances et compétences, et faisant notamment fi des dimensions contextuelles de l'évaluation ainsi que des aspects dynamiques de l'apprentissage en jeu lorsqu'un enseignant s'approprie des connaissances et tente de les mettre en pratique.

Il semble donc réducteur de considérer que l'AL ne se construit qu'en validant l'acquisition de connaissances ou la preuve de compétences une fois pour toutes, en référence à des standards universels. Tout d'abord, car maîtriser des connaissances ne signifie pas *de facto* être apte à les transférer dans la pratique (Xu & Brown, 2016). Ensuite, parce que les compétences évaluatives peuvent varier d'un contexte à l'autre, en fonction des prescriptions, des cultures, des systèmes scolaires, des équipes d'enseignants, des élèves, des disciplines, des établissements ou encore des curricula. Enfin, car le développement de compétences en évaluation demande du temps, de multiples occasions de se confronter aux expériences de la pratique et d'éprouver les connaissances à la réalité des contextes (McMillan, 2003). En effet, les enseignants ont leur propre manière de comprendre et de questionner leurs pratiques. Ce sont leurs interrogations et problèmes qui vont conditionner leur manière de les penser, voire de les infléchir (Brookhart, 2002). Il est dès lors délicat d'uniformiser ces processus d'appropriation.

Pourtant, Mertler (2009), dans un texte fréquemment cité, explicite une démarche de formation à l'AL menée auprès d'enseignants du primaire de différents établissements. Il montre les bénéfices de mesurer (*measure*) les connaissances et compétences des participants à l'aide de

⁶⁴ L'auteur propose par exemple un questionnaire en ligne pour que tout enseignant intéressé puisse mesurer son degré de développement en AL : <http://pareonline.net/htm/v8n22/cali.htm>.

questionnaires avant et après la formation relativement aux standards. L'outil qu'il utilise et qu'il a créé à ces fins est l'*Assessment Literacy Inventory* (Mertler & Campbell, 2005), un répertoire de huit tâches (*tasks*) d'évaluation emblématiques que tout enseignant devrait être capable de réaliser. Il montre qu'après deux semaines de formation, tous les enseignants ont augmenté notablement leurs scores sur chaque tâche. Il infère de ces résultats un développement de compétences et une meilleure connaissance générale des phénomènes évaluatifs chez les enseignants. Mais que signifie mesurer une compétence évaluative ? Que dit une telle démarche par questionnaire de la conception de la compétence, voire de son évaluation ? À quelles conditions pourrait-on entrevoir le développement de compétences évaluatives comme un processus complexe, inhérent aux individus et à leur questionnement saisi dans des contextes chaque fois spécifiques ?

2.2.3 Une seconde épistémologie de l'*assessment literacy* plus dynamique

D'autres chercheurs inscrivent l'AL dans une épistémologie différente, référée aux acteurs et à leurs problématiques, aux contextes, aux dimensions sociales et éthiques de leurs pratiques évaluatives. Ainsi, Willis, Adie et Klenowski (2013) considèrent l'AL comme un enjeu fondamental de la profession enseignante qu'il faut urgemment mettre en débat afin de mieux l'explicitier. Pour elles, la perspective théorique doit être repensée pour dépasser le seul statut d'outil de mesure. Pour cela, une nouvelle problématisation de l'AL, construite avec les enseignants pour en dégager une compréhension commune, est à envisager. Cette problématisation articule ce que la recherche a mis en évidence en termes de connaissances et compétences clés en évaluation avec les questions émergeant des pratiques⁶⁵ (Xu & Brown, 2016). C'est ce qui les amène à définir l'AL comme « a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students » (p. 242). La dimension dynamique de l'AL se retrouve ici à plusieurs niveaux :

- Les pratiques évaluatives au sens large du terme s'inscrivent et s'appréhendent dans des contextes sociaux ;

⁶⁵ Ou « questions vives de la pratique », comme nous le proposons dans notre conclusion de manuscrit.

- Les connaissances disciplinaires à évaluer sont situées culturellement et font l'objet de négociations permanentes de manière à ce que les élèves aient les moyens d'atteindre les objectifs d'apprentissage visés ;
- L'évaluation est intégrée aux processus d'enseignement et d'apprentissage, toujours contextualisés.

Le contexte évaluatif est dès lors décrit ici par les auteures comme complexe, puisqu'il est constitué de plusieurs dimensions interdépendantes (*layered*). Dans ce sens, la formation à l'AL valorise les débats entre pairs au travers de confrontations de pratiques évaluatives articulant création et analyse d'épreuves⁶⁶, afin de « give participants the opportunity to read and critically problematise and recontextualise norms of assessment communities and understand assessment demands » (Willis et al., 2013, p. 252)⁶⁷. Dans leur conceptualisation, Willis et ses collègues insistent sur le fait que l'AL doit, *in fine*, servir les apprentissages des élèves. C'est selon nous une dimension contextuelle supplémentaire qui montre combien leur orientation épistémologique est différente des chercheurs considérant la perspective théorique comme relativement statique.

D'autres apports significatifs sur cette conceptualisation de l'AL émergent chez Edwards (2017). Dans une recherche menée auprès d'enseignants et d'étudiants du secondaire en sciences, elle tente de montrer les dynamiques en jeu dans le développement de leur AL en évaluation sommative. Pour cela, elle se réfère à des critères qu'elle nomme *Summative Assessment Literacy Rubric* (SALRubric) permettant de suivre ce développement, et non de le mesurer. Il est intéressant de noter que leur conceptualisation suit le même mouvement que celui opéré par Willis et ses collègues, à savoir une articulation entre les connaissances issues de la recherche sur les compétences évaluatives et les questions du terrain⁶⁸. Ainsi, ses critères ont fait l'objet de reformulations, d'affinements, afin de donner à voir les connaissances et compétences clés à développer chez les enseignants dans le temps. Cette démarche explicite selon nous plus précisément certaines caractéristiques de l'AL saisies au cœur d'une telle épistémologie :

⁶⁶ Ce postulat épouse parfaitement les caractéristiques de notre dispositif de recherche-formation, ce qui représente à nos yeux un gage de validité.

⁶⁷ Nous retrouvons ici les caractéristiques de la modération sociale qui désigne « des pratiques collectives de confrontation de « jugements humains » sur des travaux d'élèves » (Mottier Lopez, Tessaro, Dechamboux & Morales Villabona, 2012, p. 4).

⁶⁸ Relevons également que cette conception est pleinement cohérente avec les postulats de la recherche collaborative au sens des chercheurs que nous convoquons dans notre partie méthodologique.

- Les connaissances en jeu dans l'AL dépendent du contexte des enseignants dans lequel elles sont potentiellement mises en œuvre sous forme de compétences ;
- La compréhension et la pratique de l'AL sont influencées par les prescriptions formulées par les décideurs, les lois, l'histoire de l'évaluation au sein des systèmes concernés et les habitudes des pratiques au sein des équipes ;
- En conséquence, les listes de compétences évaluatives deviennent plus pertinentes lorsqu'elles sont lues au travers des contextes dans lesquels l'évaluation est mise en œuvre (Allal, 2009 ; Braxmeyer et al., 2005 ; Jorro, 2016 ; Mottier Lopez, 2015 ; Shepard, 2000 ; Spillane & Miele, 2007).

La question de départ à laquelle l'AL tente de répondre change donc, et pourrait se formuler ainsi : « à quels niveaux de leurs pratiques évaluatives les enseignants rendent-ils compte de tensions, de manques, de problèmes et quelles connaissances et/ou compétences leur permettrait potentiellement de les résoudre ? ». Ce renversement n'est pas anodin. Il met au centre de la conceptualisation de l'AL la nécessité de prendre en compte les compromis (*compromises*) que les enseignants font en permanence quand ils évaluent, et qui sont de nature profondément dynamique également (Carless, 2011) puisqu'ils traduisent des processus de décision perpétuellement influencés par le contexte. Dans ce sens, Xu et Brown (2016) parlent plutôt de *teacher assessment literacy in practice* (TALiP), une perspective théorique d'AL faisant l'objet de négociation contextualisées entre les questions des enseignants et leur conception de l'évaluation, les éléments contextuels dans lesquels ils évoluent et les attentes en termes de connaissances et compétences en évaluation à développer. Comme le disent ces derniers chercheurs :

Obviously TALiP is neither a static concept nor an idealized body of knowledge and skills. Rather, it is a dynamic, complex entity combining teachers' assessment knowledge, their conceptions of assessment, and their responses to the external contexts embedded with actual constraints and affordances in the environment. TALiP reflects the true realities concerning how teacher AL is enacted in the classroom. The advancement of one's TALiP is not straightforward; it needs to be driven by principles of teacher learning. (Xu & Brown, 2016, p. 23)

Dans cette perspective, l'enseignant n'est plus « *literate* » ou « *illiterate* ». Le développement de l'AL suit ici des étapes non linéaires, parfois imprévisibles, complexes, à multiples dimensions (culturelles, sociales, cognitives, éthiques, environnementales). Développer des connaissances et des compétences dans une telle épistémologie se réalise dans un continuum,

en acceptant que différents niveaux de maîtrise se côtoient chez l'enseignant, que des acquis puissent participer au développement de plusieurs compétences, ou encore que des lacunes persistent et ainsi provoquent des ruptures. Nous faisons par ailleurs l'hypothèse que ce développement est également dépendant de la manière avec laquelle la formation à l'AL est dispensée, tant dans sa forme (durée, ingénierie) que dans son fond (inductive, déductive, abductive, avec ou sans modèle théorique).

Ainsi, Xu et Brown (2016) proposent une nouvelle définition de l'AL inscrite dans leur modèle. Pour eux, la TALiP consiste en trois niveaux de maîtrise :

- Une maîtrise des connaissances de base en évaluation, qui incluent les principes fondamentaux du « quoi », du « pourquoi » et du « comment », sans lesquels les enseignants ne peuvent pas penser leurs pratiques évaluatives plus profondément ;
- En référence à ces connaissances, une compréhension approfondie de ce que devrait être concrètement une démarche évaluative, tenant compte des processus d'enseignement et d'apprentissage, et articulant connaissances théoriques et questions de la pratique ;
- Une capacité à traduire les prescriptions en évaluation de manière à les adapter à la classe.

Ces trois niveaux - qui selon nous articulent de manière intéressante pour notre problématique de recherche la question des connaissances, de leur mobilisation sous forme de compétences et de leur inscription dans un contexte - traduisent une « capacité d'évaluation »⁶⁹ de l'enseignant (Volante & Fazio, 2007) en construction permanente. L'enseignant évaluateur compétent est donc ici un professionnel en réflexion constante sur ses pratiques, qui se forme avec ses pairs, des formateurs et/ou des chercheurs, engagé dans des discussions sur l'évaluation et des projets en rapport, en recherche de ressources pour améliorer ses pratiques et mieux comprendre son rôle d'évaluateur. Toute recherche, comme la nôtre, visant à investiguer le développement de l'AL peut entrer par l'un ou l'autre des niveaux du TALiP. C'est la raison pour laquelle il est important maintenant d'approfondir la question des connaissances et des compétences en jeu au sein de cette perspective théorique.

⁶⁹ Ce terme est la traduction littérale d'*assessment literacy* que les auteurs utilisent dans leur texte. Nous parlerons quant à nous de compétences évaluatives sommatives.

2.3 De quelles connaissances parle-t-on dans une perspective théorique d'*assessment literacy* dynamique ?

Dans une revue de littérature portant sur l'AL depuis les années 90, Xu et Brown (2016) ont tenté de recenser les connaissances jugées importantes au cœur des différentes conceptions recensées. Dix-sept recherches ont été analysées. Il en ressort une variété importante de connaissances qu'il serait trop long de lister ici. Notons toutefois qu'elles font toutes référence, plus ou moins explicitement, à la nécessité, pour tout enseignant, de maîtriser des connaissances fondamentales et transversales en évaluation⁷⁰, de comprendre les standards⁷¹, et de décrire son contexte.

Un consensus fin à ce niveau n'est donc pas possible, ni par ailleurs souhaitable. En effet, comme nous l'avons vu, dans une perspective dynamique, les connaissances de base à acquérir pour un développement en AL sont définies au regard des questions des enseignants, des résultats de recherches sur les pratiques évaluatives sommatives, des contraintes contextuelles et des enjeux de la pratique relativement aux populations d'élèves concernées. Toutefois, en poursuivant cette démarche, Xu et Brown (2016) identifient tout de même sept catégories de connaissances indispensables (*knowledge base*) à un développement en AL (pp. 19-20), dont une grande partie font sens dans le cadre de notre recherche. Ce sont des connaissances :

- Disciplinaires sur les objets de savoir et relatives aux contenus des curricula ;
- Des buts, méthodes et contenus des évaluations, notamment en ce qui concerne les tâches évaluatives ;
- En pondération (exploitation de divers systèmes) et en notation ;
- En feedback ;
- En évaluation mutuelle et en auto-évaluation ;
- Au niveau de l'interprétation et de la communication des résultats ;
- En éthique de l'évaluation.

Ces catégories de connaissances représentent un échantillon général de connaissances clés devant s'inscrire dans des orientations théoriques explicites. Elles sont nécessaires au développement de compétences évaluatives, car elles vont en être les descripteurs. Mais elles

⁷⁰ Popham (2011) exemplifie ces connaissances avec les concepts de « validité » et de « biais évaluatifs » (*test bias*), que tout évaluateur devrait, selon lui, maîtriser. Mais il ne va plus loin.

⁷¹ Les standards peuvent être assimilés, dans notre contexte, aux éléments prescriptifs présents dans un référentiel d'enseignement et d'apprentissage, comme le plan d'études romand (PER), ou dans un référentiel d'évaluation lorsqu'il existe.

nécessitent d'être contextualisées. Par exemple, il s'agira, pour la troisième catégorie, de travailler l'appropriation de connaissances relatives à l'évaluation critériée et critérielle en fonction des connaissances de la recherche, des questions des enseignants et de leurs pratiques réelles en la matière, ainsi qu'au regard des prescriptions institutionnelles. C'est en cela que ces connaissances sont dynamiques, puisque leur développement sous forme de compétences ne peut se réaliser qu'au travers d'un filtre chaque fois spécifique.

Cela étant, ces catégories de connaissances restent insuffisantes si elles ne sont considérées que pour elles-mêmes. En effet, dans une perspective d'AL dynamique, il est délicat de considérer que l'acquisition de connaissances garantit un développement de compétences évaluatives. Pour tout enseignant, le processus de mise en œuvre ou de mobilisation de connaissances dans des situations d'évaluation réelles tient plutôt du compromis que du transfert, les connaissances entrant systématiquement en tensions avec les conceptions de l'évaluateur, son contexte, les prescriptions, et bien entendu potentiellement avec les modalités de formation offertes. Et c'est bien ce compromis qui conditionne la nature des compétences clés à évaluer que les enseignants vont être amenés à développer de manière continue dans une perspective d'AL.

2.4 Qu'en est-il des compétences à évaluer en *assessment literacy* du point de vue de la recherche ?

Notre conceptualisation de l'AL étant dynamique, nous ne nous inspirerons ici que des travaux de chercheurs qui ont tenté d'élaborer des compétences à évaluer s'inscrivant dans cette épistémologie.

2.4.1 Un premier ensemble de réflexions intéressantes pour une *assessment literacy* dynamique : les travaux de Brookhart (2011)

Cette chercheuse, en se référant aux standards nationaux décrivant les compétences professionnelles attendues chez les enseignants⁷², en énonce 11, participant au développement de l'AL. Quatre d'entre elles sont à nos yeux pertinentes pour notre problématique :

...

- II Teachers should be able to articulate clear learning intentions that are congruent with both the content and depth of thinking implied by standards and curriculum goals, in such a way that they

⁷² The 1990 Standards for Teacher Competence in Educational Assessment of Students, Etats-Unis.

are attainable and assessable ;

...

IV Teachers should understand the purposes and uses of the range of available assessment options and be skilled in using them ;

V Teachers should have the skills to analyze classroom questions, test items and performance assessment tasks to ascertain the specific knowledge and thinking skills required for students to do them ;

...

VII Teachers should be able to construct scoring schemes that quantify student performance on classroom assessments into useful information for decisions about students, classrooms, schools, and districts. These decisions should lead to improved student learning, growth, or development.

... (p. 7)

Plus spécifiquement, nous observons que ces compétences s'appuient sur des connaissances présentes au sein de notre modèle théorique de l'alignement curriculaire élargi. En effet, la II fait référence au curriculum, aux objectifs et aux contenus du plan d'études. La IV met en évidence l'expertise à avoir au niveau de la création des items et des tâches, référencées à la même taxonomie que celle que nous avons mise en œuvre dans notre dispositif (Anderson & Krathwohl, 2001). La V explicite la nécessaire articulation qui doit exister entre les tâches évaluatives et les objectifs qu'elles sont censées évaluer. Et la VII met en dialogue les processus de pondération et de notation.

Il est donc intéressant de noter qu'à ce titre, certaines de ces compétences, en forte cohérence avec notre modèle théorique, sont très inspirantes pour élaborer ce que serait notre « référentiel » de compétences évaluatives.

2.4.2 Un second apport significatif sur une *assessment literacy* dynamique : les recherches de Popham (2009)

Pour ce chercheur, l'enjeu premier de l'AL est de permettre aux enseignants de prendre de meilleures décisions. Cette perspective théorique est pour lui un outil de développement professionnel. Tout programme de formation continue à l'AL poursuivant ce but devrait dès lors se concentrer sur quelques compétences clés articulées à des connaissances et s'appuyant

sur des contenus. Il en formule treize. Quatre d'entre elles retiennent tout particulièrement notre attention :

...

3. The prominent role three types of validity evidence should play in the building of arguments to support the accuracy of test-based interpretations about students, namely, content-related, criterionrelated, and construct-related evidence.

...

5. Construction and improvement of selected response and constructed-response test items.

6. Scoring of students' responses to constructed-response tests items, especially the distinctive contribution made by wellformed rubrics.

...

10. Interpreting students' performances on large-scale, standardized achievement and aptitude assessments.

... (pp. 8-10)

À nouveau, ces compétences sont éclairantes pour nous. Nous observons que la 3 définit la validité en termes de rapport au contenu évalué. La 5 mentionne la capacité à construire des items et tâches évaluatives affûtées (*honing*), c'est-à-dire précises du point de vue des objets de savoir sur lesquelles elles portent. La 6 spécifie que tout enseignant devrait savoir comment construire et exploiter des critères, et plus largement être apte à choisir le meilleur système pour mettre en évidence le poids de l'apprentissage évalué. Quant à la 10, elle explicite les connaissances nécessaires à la notation, notamment au niveau du traitement des scores des élèves et de la création d'outils de notation comme les échelles.

La recherche circonscrit donc les connaissances et compétences nécessaires à une formation à l'AL. Nous sommes dès lors outillé conceptuellement pour articuler ces apports avec notre contexte spécifique, qui se veut à deux dimensions : une première relative aux contextes des enseignants dans lesquels ils sont amenés à mettre en œuvre leurs pratiques évaluatives sommatives ; et une autre relative à notre dispositif de formation. En effet, ce sont bien ces deux dimensions qui vont infléchir les processus d'appropriation du modèle théorique et de développement des compétences évaluatives des enseignants dans leur pratique.

2.5 Les compétences en *assessment literacy* retenues dans le cadre de notre recherche

En référence à nos questions de recherche en fin de cette partie, en articulant les différentes composantes de notre modèle théorique, les apports de la recherche et les dimensions

contextuelles mentionnées plus haut, nous nous proposons de poursuivre notre travail du point de vue de l'AL en formulant six compétences évaluatives, comme autant de références :

- A. Analyser le plan d'études du point de vue des objectifs et des objets de savoir disciplinaires qui y circulent à des fins d'évaluation sommative en mettant en relation de manière cohérente les différents niveaux de formalisation des objectifs, et en analysant les objets de savoir pour en déterminer les caractéristiques essentielles à des fins d'évaluation sommative.
- B. Elaborer des tâches évaluatives complexes en rapport avec les objectifs des plans d'études qui donnent des preuves d'apprentissages maîtrisés ou en voie de réalisation, en privilégiant des tâches qui donnent à voir des unités d'apprentissage significatives (approche qualitative).
- C. Adopter des systèmes de pondération permettant de mettre en évidence le poids donné à l'apprentissage référé aux objectifs et aux objets de savoir en privilégiant, tant que possible, une pondération critériée, notamment pour permettre un feedback aux élèves à visée de régulation.
- D. Construire des notes de manière à ce qu'elles rendent compte des acquis et lacunes des apprentissages des élèves en élaborant ses propres échelles et barèmes en référence au système de pondération, à la complexité des tâches et aux objectifs évalués.
- E. Penser ses pratiques évaluatives sommatives en termes de validité et de cohérence d'alignement curriculaire élargi en les référant aux processus d'enseignement et d'apprentissage, et en analysant leurs cohérences à l'aune d'éléments taxonomiques.
- F. Justifier ses choix en matière de pratiques évaluatives sommatives au regard de son contexte institutionnel cantonal, local (établissement) et de la culture des équipes dans lesquelles elles s'inscrivent, en les adaptant aux prescriptions institutionnelles, tout en étant attentif à ce qu'elles soutiennent les apprentissages des élèves.

La formulation et le contenu sur lequel portent ces compétences tiennent compte de la réalité des enseignants. Comme le disent Xu et Brown (2016) : « teachers who are placed in complex contexts have to make professional decisions about assessment in response to various factors that may facilitate or inhibit their practice » (p. 17)⁷³. L'enjeu est donc ici ni plus ni moins de construire une perspective théorique d'AL réaliste, soutenant, permettant au plus grand nombre de progresser dans ses pratiques évaluatives sommatives.

⁷³ Certaines propositions de la littérature scientifique francophone sur le jugement professionnel que nous présentons dans le cadre conceptuel s'inscrivent dans cette perspective.

Les pistes les plus fécondes pour éprouver la validité d'une perspective théorique d'AL comme la nôtre devraient idéalement se développer au travers de dispositifs de formation à l'évaluation en équipe (Willis et al., 2013), afin de partager un langage commun et de négocier certains choix évaluatifs (Fleer, 2015 ; Lukin et al., 2004). Les actions de formation, spécifiquement, devraient porter sur des démarches d'évaluation, comme la création de critères, le *design* de tâches évaluatives (Adie, 2013). Un temps conséquent devrait y être alloué, de manière à ce que le développement des compétences puisse se réaliser de manière progressive et s'ancrer dans des expériences de pratiques en classe (Lukin et al., 2004). Autant de caractéristiques que nous avons respectées dans le dispositif de recherche-formation concerné par notre recherche.

3 Pour ou contre une *assessment literacy* en évaluation sommative ?

Nous l'avons vu, conceptuellement, l'AL vise notamment à théoriser une compréhension approfondie de ce qu'est l'évaluation scolaire telle qu'elle devrait être menée par les enseignants pour qu'ils prennent les meilleures décisions pour les élèves. La pratique évaluative, saisie dans l'acception dynamique de cette perspective théorique, est à considérer aussi bien du point de vue diagnostique, formatif ou sommatif (Volante & Fazio, 2007). Par ailleurs, l'évaluation entretient un rapport de cohérence avec les processus d'enseignement et d'apprentissage, dont la connaissance des objets de savoir et des curricula en est le fil rouge. Willis et ses collègues (2013) vont même jusqu'à affirmer que « the assessment literacy [...] needs to be positioned within a discussion about learning theory » (p. 252). Elles défendent également l'idée que les pratiques évaluatives se conjuguent aux pratiques pédagogiques et didactiques lorsqu'il s'agit de soutenir l'apprentissage des élèves, un postulat auxquels adhèrent également bon nombre de chercheuses et chercheurs francophones (e. g. Allal & Mottier Lopez, 2007 ; Amigues & Zerbato-Poudou, 1996 ; Durand & Chouinard, 2012 ; Laveault, 2004 ; Mottier Lopez, 2010, 2015 ; Mottier Lopez & Laveault, 2008 ; Timperley, 2011 ; Vial, 2012).

Ce positionnement théorique se traduit notamment dans les référentiels de compétences en AL qui ne se concentrent pas sur une seule fonction de l'évaluation. Ce sont seulement dans des recherches étudiant précisément certaines pratiques particulières d'évaluation, comme la nôtre, que nous voyons apparaître une *assessment literacy* en évaluation sommative (e. g., DeLuca & Klinger, 2010 ; Edwards, 2017 ; Klinger, Volante & DeLuca, 2009). Toutefois, nous parlerons dans notre cas toujours d'*assessment literacy*, sans le référer spécifiquement à l'évaluation sommative, pour deux raisons essentielles :

- Les compétences évaluatives que nous avons construites touchent des dimensions périphériques à celles qui se jouent a priori en évaluation sommative ;
- Notre modèle théorique qui conditionne le raisonnement conceptuel de l'AL s'inscrit pleinement en cohérence avec les processus d'enseignement et d'apprentissage.

4 En guise de synthèse

« There is still relatively little research devoted to understanding the assessment literacy of classroom teachers » (Volante & Fazio, 2007, p. 750)⁷⁴. Par ces mots, ces deux chercheurs canadiens lancent un plaidoyer pour que différentes perspectives théoriques d'AL soient mises à l'épreuve de la réalité des classes.

Notre recherche s'inscrit pleinement dans cet appel. Tout d'abord au niveau méthodologique, où notre méthode collaborative décrite dans la partie 6, inscrite dans une perspective compréhensive et interprétative, tente de saisir à quelles conditions les problèmes et questions soulevés par les enseignants peuvent être pris en compte et dépassés lorsque ces derniers sont engagés dans un processus de formation à l'évaluation sommative. Au niveau conceptuel ensuite, au travers du modèle théorique de l'alignement curriculaire élargi sur lequel elle s'appuie pour donner à voir les phénomènes évaluatifs que ce dernier va pouvoir éclairer : nous faisons l'hypothèse que ses composantes raisonnent avec les connaissances que les enseignants devraient posséder pour développer leurs compétences évaluatives sommatives. Et enfin d'un point de vue des apports qu'elle pourrait être en mesure de mettre en évidence, tant pour la recherche que pour la formation.

À la lumière de ces éléments conceptuels, nous pouvons formuler nos questions de recherche.

⁷⁴ Constat confirmé par Christopher DeLuca (2018, correspondance personnelle).

5 Questions de recherche

Les deux premières questions générales de recherche sont les suivantes :

- A Sur quelles dimensions de leurs pratiques évaluatives les enseignants portent-ils leur attention lorsqu'elles sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi ?

Le but ici est d'identifier les dimensions retenues par l'enseignant, de mettre en évidence à quel.s niveau.x se situent les problèmes qu'il exprime, et de montrer comment ces derniers évoluent lorsque le praticien s'approprie le modèle théorique de l'alignement curriculaire élargi. Par exemple, quand il questionne ce qu'est une évaluation « juste », que devient ce terme mis à l'épreuve du modèle théorique ?

Cette question générale se divise ainsi en trois sous-questions :

- A1. Sur quelles dimensions de leurs pratiques évaluatives les enseignants portent-ils leur attention avant la formation ?

- A2. Quels points de cohérence et de rupture d'alignement les enseignants, avec le guidage du formateur, observent-ils dans leurs pratiques d'évaluation sommative lorsqu'ils se réfèrent au modèle de l'alignement curriculaire élargi ?

- A3. Quelles composantes de l'alignement curriculaire élargi semblent permettre aux enseignants d'approfondir leurs analyses des ruptures au fil du temps ?

La deuxième question générale de recherche porte sur le développement hypothétique des compétences évaluatives des enseignants.

- B Quelles compétences évaluatives spécifiques d'un point de vue sommatif l'alignement curriculaire élargi participe-t-il à développer ?

Le but ici est de problématiser le développement des compétences en *assessment literacy* dont nous faisons l'hypothèse qu'elles se développent au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi.

Cette question générale se divise en deux sous-questions :

B1. Sur les six compétences évaluatives considérées, lesquelles se développent principalement ?

B2. Relativement aux pratiques effectives des enseignants, en quoi le développement des compétences évaluatives rend-il compte d'une épistémologie dynamique de l'*assessment literacy* ?

Comme annoncé plus haut, la troisième question générale de recherche cible les pratiques de notation.

C Comment les enseignants décrivent-ils leurs pratiques de notation ?

Le but ici est de donner à voir sur quels phénomènes relatifs à la notation les enseignants portent leur attention, soit pour manifester des problèmes rencontrés, soit pour expliciter des questions professionnelles nécessitant des apports.

Cette question générale se divise en trois sous-questions :

C1 : Sur quel.s phénomène.s et/ou question.s les enseignants portent-ils leur attention relativement à leurs pratiques de notation avant la formation ?

C2 : Quelle est la nature des obstacles rencontrés par les enseignants dans leurs pratiques de notation ?

C3 : Quels apports ces résultats offrent-ils sur les pratiques de notation du point de vue de l'alignement curriculaire « élargi » ?

Partie 5 - Présentation du dispositif de recherche

1 Introduction

Nous exposerons ici des éléments contextuels de notre recherche. Il nous semble important de commencer par présenter les phases du recrutement des enseignants ayant collaboré à notre projet, car elles donnent à voir les difficultés rencontrées à trouver huit volontaires et par là, interrogent selon nous l'ouverture du système scolaire à la formation à l'évaluation.

Puis, nous ferons état des profils des participants. Ils sont importants pour mieux appréhender ces enseignants au travers de leur parcours professionnel, leur âge, leur formation, leur contexte et leurs diverses expériences professionnelles.

Nous poursuivrons en présentant dans le détail le déroulement de la recherche-formation mise en oeuvre. Il permettra de mieux comprendre les différentes phases dans lesquelles nous avons engagé les participants, leurs finalités et articulations.

Nous clôturerons brièvement en décrivant l'entrée en matière avec les participants, de manière à expliciter les premiers échanges, fort importants pour saisir notre orientation méthodologique.

2 Contexte de la recherche

2.1 Une première stratégie de recrutement d'enseignants infructueuse

Au départ, nous avons eu le projet d'engager dans le processus de recherche-formation huit enseignants du secondaire, quatre en français, et quatre en sciences humaines (histoire, géographie, économie et citoyenneté). Notre but étant de mener une recherche qualitative centrée sur l'analyse de pratiques évaluatives, cet effectif nous semblait suffisant.

D'emblée, nous n'avons pas souhaité collaborer avec des personnes ayant déjà suivi une formation continue à l'évaluation, afin de partir de pratiques courantes, traditionnelles. Notre choix de disciplines s'est fait en fonction de leur importance dans la scolarité et des enjeux qu'elles sous-tendent en matière d'évaluation sommative : le français à enjeu fort et à poids conséquent, les sciences humaines à enjeu modéré et à poids restreint⁷⁵. À ce moment-là, il

⁷⁵ Au secondaire, le français fait partie du groupe de disciplines I avec les mathématiques, l'allemand, les sciences de la nature et l'option spécifique, alors que l'histoire et la géographie font partie du groupe II, avec l'anglais. Le groupe de disciplines III est constitué des arts visuels, de la musique, des activités créatrices et manuelles ou de l'éducation nutritionnelle.

nous a en effet semblé pertinent de comparer les caractéristiques des pratiques évaluatives sommatives au sein de ces deux disciplines. Nous avons alors commencé à solliciter nos relations dans divers établissements à l'aide du document présenté en annexe Partie 7, 1, afin d'entrer en contact avec de potentiels volontaires. Ce document visait à expliciter les buts du projet, mais aussi à délimiter le travail demandé aux enseignants qui s'y investiraient.

Nous avons envoyé ce cadre de travail à dix-neuf personnes : cinq chefs d'établissement (trois anciens collègues, une connaissance, et un rencontré lors d'une formation continue), huit doyens (un ancien collègue, un ancien étudiant, deux connaissances, deux ayant participé à une formation continue et deux recommandés par un directeur), trois Praticiens Formateurs⁷⁶ ayant également suivi une formation continue sur cet objet avec nous et deux enseignants recommandés par leur directeur. Cette première salve de demandes a touché quatorze établissements, situés dans toutes les régions, milieux socio-économiques et contextes du canton de Vaud. Nous avons communiqué avec nos interlocuteurs essentiellement par email et par téléphone, en devant fréquemment les relancer suite à leur silence. Malgré cette insistance, aucune de nos demandes n'a abouti. Sur les cinq chefs d'établissement, seul deux nous ont dit avoir relayé la demande à leur corps enseignant lors de conférences ou de séances de travail. Les trois autres, s'ils ont souligné l'intérêt de la recherche, ont tous fait état d'une surcharge de travail des enseignants les empêchant de s'engager dans le projet. Du côté des doyens, cinq nous ont affirmé avoir tenté de démarcher du côté de leurs collègues, selon des modalités qu'ils ne nous ont pas explicitées. Un enseignant de langues s'est ainsi manifesté, motivé par le fait qu'il nous connaissait d'un cadre privé. Les Praticiens Formateurs, quant à eux, nous ont dit avoir également tenté d'approcher des collègues en leur vantant les mérites de cette formation. Sans succès. Restait alors un enseignant, chef de file de français dans son établissement. Nous avons plusieurs fois essayé de trouver un moment pour rencontrer ses collègues, mais leur emploi du temps et leurs plans de séances étaient si denses que nous n'y sommes jamais arrivés. Nous avions prévu de commencer la formation en août 2014. Au mois de mai, nous n'avions qu'une seule personne intéressée. Nous avons donc revu notre stratégie de recrutement.

2.2 Une nécessaire réorientation du recrutement

Nous avons alors contacté deux enseignants ayant participé à une journée d'étude sur

⁷⁶ Le Praticien Formateur est un formateur de terrain, formé par la HEP Vaud, qui accompagne les étudiants en stage.

l'évaluation en mai 2013, dans laquelle nous sommes intervenu⁷⁷. Le premier, spécialiste de mathématiques, et très intéressé par le projet, nous a mis en contact avec deux de ses collègues, un spécialiste d'histoire, et l'autre d'économie. Une séance nous a permis d'échanger sur les questions, problèmes et besoins qu'ils avaient. Là, nous avons constaté que les sciences humaines et sociales (SHS), telles que présentées notamment dans le Plan d'Etudes Romand (PER), regroupent des disciplines fort différentes⁷⁸, et donc des questions relatives à l'évaluation sommative très spécifiques. À l'issue de cette séance, nous avons donc pris la décision de nous centrer sur une autre discipline phare : les mathématiques. Et ceci, même si les deux enseignants en question nous ont manifesté leur motivation à participer, contrairement à celui de mathématiques, qui a renoncé, son emploi du temps étant trop lourd.

2.3 Des changements importants dans le choix des disciplines

Nous avons donc dû encore une fois revoir notre recherche de volontaires. Dès le mois de juin 2014, nous avons poursuivi nos démarches intensivement, mais en nous concentrant cette fois sur le français et les mathématiques. Si ce choix s'est réorienté en raison de contraintes pratiques, il nous a rapidement semblé qu'aborder deux disciplines, et non plus sept (!), et de surcroît à forts enjeux, rendrait la recherche plus dense, tout en autorisant toujours, au besoin, une perspective comparative.

Concernant les mathématiques, des circonstances favorables nous ont permis d'entrer en contact avec un établissement ayant fait recours à nos services pour une action de formation sur la notation. Là, lors d'une séance d'introduction en présence de toute la file⁷⁹, nous avons présenté notre projet, et trois enseignants sur quatorze, que nous présentons plus loin, nous ont dit vouloir y participer. Nous les appellerons Philémon, Catherine et Paul⁸⁰. Parallèlement, et après avoir sollicité l'autorisation de la Direction Générale de l'Enseignement Obligatoire (DGEO), nous avons diffusé notre recherche au travers du réseau educanet⁸¹ et son groupe thématique des mathématiques, véritable plateforme d'échange et de communication de tous

⁷⁷ Les pratiques des enseignants en évaluation : que nous apprennent-elles ?, IRDP, Neuchâtel, 31 mai 2013.

⁷⁸ Les disciplines obligatoires de ce domaine sont l'histoire, la géographie et la citoyenneté, auxquelles s'ajoutent l'économie, le droit ainsi que l'éthique et les cultures religieuses en disciplines optionnelles.

⁷⁹ Une file rassemble tous les enseignants d'un établissement qui enseignent la même discipline dans des degrés similaires.

⁸⁰ Prénoms d'emprunt, pour tous les enseignants cités dans ce manuscrit.

⁸¹ Avec ses 638'000 membres, educanet² est la principale plateforme d'enseignement et d'apprentissage en ligne de Suisse. Elle offre aux enseignants vaudois des ressources multiples, notamment au travers de groupes thématiques dont la Direction Générale de l'Enseignement Obligatoire (DGEO) est la référence (<https://www.educanet2.ch/>).

les enseignants de la discipline du canton. Touchant plusieurs centaines de personnes, cette démarche nous a permis de rencontrer cinq spécialistes motivés ayant répondu à l'appel. Là, suite à des entretiens individuels, nous avons choisi le quatrième volontaire du groupe, Fernand, l'établissement dans lequel il enseignait ayant les mêmes caractéristiques socio-économiques que celles de ses collègues.

Conjointement, nous avons persévéré dans le recrutement des quatre volontaires pour le français. La rencontre d'Anouk, qui était également sur la liste des participants à la journée d'études sur l'évaluation et que nous avons contactée à la même période, a été un succès. Très motivée et en questionnement, elle s'est tout de suite engagée. Elle a rapidement parlé du projet à sa collègue directe, Alizée, que nous avons vite rencontrée et qui nous a donné son aval pour une participation. Grâce à elles, et par un effet de ricochet, nous sommes ensuite entré en contact avec Samia, également enseignante dans le même établissement, puis enfin avec Julie, toujours collègue directe, qui ont toutes deux également accepté de rejoindre le groupe⁸². Au mois de septembre 2014, l'équipe de huit volontaires était - enfin - constituée.

2.4 Une campagne de recrutement difficile

Les démarches, depuis notre première demande de contact jusqu'à l'accord de la dernière participante, ont pris plus de six mois. Nous pensions que ce recrutement serait long. Mais nous avons tout de même été surpris de rencontrer autant d'obstacles. Cet état de fait interpelle et questionne. En effet, dans quelle mesure cette difficulté ne cacherait-elle pas d'autres problèmes, contextuels, idéologiques ou relatifs aux pratiques que les acteurs ne souhaiteraient pas - ou ne pourraient pas - pour de multiples raisons, exposer ? Dans le même sens, si nous imaginons bien que les enseignants font face à une charge conséquente de travail due notamment à des changements majeurs, comment expliquer que certains d'entre eux ont répondu favorablement à notre appel alors même que leur directeur nous avait verrouillé leur accès ? Y aurait-il un risque potentiel, pour un chef d'établissement, à voir certains de ses enseignants acquérir de nouvelles connaissances et développer leurs compétences évaluatives ? Autant d'interrogations qui nous obligeront à considérer le contexte comme un élément clé infléchissant le développement des compétences à évaluer⁸³.

⁸² Afin de ne pas surcharger le texte, nous ne faisons pas état des nombreuses - et infructueuses - autres démarches informelles réalisées dans le cadre de ce recrutement.

⁸³ Ce rapport au contexte est approfondi dans le détail dans la partie méthodologique (6) et dans le cadre conceptuel sur l'*assessment literacy* (4).

3 L'entrée en matière avec les participants

Avant de réunir les huit participants au sein de notre dispositif décrit plus loin, nous avons eu un entretien avec chacun d'entre eux, sauf pour le trio de spécialistes de mathématiques issu du même établissement avec lequel nous avons débattu lors de la séance de formation demandée par leur direction. Ces premiers échanges ont permis de faire connaissance, de parler de manière informelle de leurs interrogations et problèmes relatifs à l'évaluation, et plus largement, de la profession. Nous avons profité de ces entrevues pour présenter les grandes lignes du dispositif, afin de clarifier l'engagement qu'il requerrait (participer à trois demi-journées de formation avec à la clé la création de deux évaluations sommatives en rapport avec la classe, écrire trois récits de pratique et réaliser deux entretiens semi-dirigés). Nous avons également mentionné les modalités de rétribution offertes par la HEP et répondu à toutes les questions. Aucun contrat n'a été signé. Nous sommes partis sur un accord de principe oral, laissant à chacun la possibilité de quitter le projet s'il le souhaitait⁸⁴. Les huit enseignants ont tous accepté les modalités de collaboration et se sont unanimement dit très motivés et curieux à l'idée de travailler avec nous sur cet objet. Par souci de clarté, nous les présentons par discipline.

4 Les profils des enseignants volontaires

Huit enseignants ont donc accepté de participer à la recherche, quatre en français et quatre en mathématiques.

4.1 Les quatre enseignantes de français

- Anouck : Titulaire d'un Master en lettres de l'Université de Lausanne, elle a obtenu son Master en enseignement secondaire à la HEP Vaud. Elle enseigne depuis huit ans. Elle a commencé sa carrière en Valais, où elle a été cheffe de file de sa discipline, avant d'arriver dans son établissement actuel où elle travaille depuis quatre ans. Elle occupe

⁸⁴ Notre recherche n'a pas fait l'objet, comme c'est exigé aujourd'hui, d'une soumission à la commission de l'éthique de la Faculté de psychologie et des sciences de l'éducation de l'Université de Genève. Cette commission n'existait pas. Au moment où cette thèse est défendue, un consentement écrit et signé est demandé à chaque participant, le contrat oral n'étant plus suffisant.

un poste à 80%, avec une maîtrise de classe et des heures de français en 9VP⁸⁵, et des heures de français, géographie, histoire et citoyenneté en 11 VSB⁸⁶.

- Alizée : Titulaire d'un Master en lettres de l'Université de Lausanne, elle a obtenu son titre d'enseignement secondaire au séminaire pédagogique pour l'enseignement secondaire (SPES). Elle enseigne depuis douze ans, et a débuté sa carrière dans l'établissement où elle est encore. Au moment de la recherche, elle occupe un poste à 58%, avec quatre périodes d'anglais en 11 VSB, trois périodes d'anglais et six de français en 9 VP où elle est également maîtresse de classe, et une demi-période de cheffe de file de français 9-10-11.
- Samia : Titulaire d'un brevet d'enseignement primaire de l'Ecole Normale, elle est au bénéfice d'une expérience d'enseignement de quatorze ans, dans divers contextes et degrés, avec des élèves souvent en difficulté. Elle a travaillé un an à la fondation Verdeil⁸⁷, dans les années de l'orientation du système scolaire vaudois, dans la voie secondaire à options (VSO). Elle a également travaillé en entreprise et comme indépendante en coaching scolaire pour lequel elle s'est formée. Au moment de la recherche, Samia est responsable d'un projet de classes « Tremplin » qui vise à donner aux élèves en grande difficulté du soutien scolaire intégré⁸⁸. Elle occupe un poste à 68%, avec des enseignements dans cette structure et dans deux classes de 10G niveau 2 pour le français⁸⁹.
- Julie : Titulaire d'un brevet d'enseignement primaire de l'Ecole Normale, elle est au bénéfice d'une expérience d'enseignement de vingt et un ans, dont quatre au primaire, trois en enseignement privé et quatorze au secondaire. Elle travaille à 95%, a une

⁸⁵ Pour une meilleure compréhension des acronymes relatifs à l'organisation du système scolaire vaudois, le lecteur se référera à l'annexe Partie 1, 1 et à l'annexe Partie 2, 2 pour une comparaison entre le système en vigueur avec la LEO et l'antécédent.

⁸⁶ Suite à sa participation à la recherche, Anouck s'est engagée comme Praticienne Formatrice à la HEP Vaud et comme formatrice auxiliaire pour donner des séminaires de formation initiale, notamment en évaluation des apprentissages. Elle a encouragé ses collègues de file à participer à une formation continue en évaluation menée par nos soins en 2016-2017 avec 13 enseignantes. Elle a ensuite pris le poste de doyenne pédagogique dans son établissement en janvier 2018.

⁸⁷ <http://www.verdeil.ch>.

⁸⁸ Projet pour les élèves de VG en grande difficulté scolaire : classe ouverte deux matins par semaine (une fois pour les élèves de 9VG, l'autre pour ceux de 10VG) ; l'élève passe une matinée par semaine dans cette structure (sur un temps déterminé de cas en cas). Le travail est principalement scolaire dans l'idée de redonner confiance aux élèves et de leur apprendre à apprendre. Toutes les périodes sont données en co-enseignement. La décision de placement à la classe Tremplin est prise d'entente entre les maîtres de classe, la Direction d'établissement, les parents et les enseignants responsables du projet.

⁸⁹ En 2017, Samia a commencé un Master en enseignement spécialisé.

maîtrise de classe en 11 VSO où elle enseigne l'allemand, le français, l'histoire, la géographie, la citoyenneté et l'approche du monde professionnel, l'enseignement de l'allemand et du français en 10G niveau 2 et l'enseignement du français dans une autre classe de 10G niveau 1. Julie a été également amenée à travailler en entreprise, dans le journalisme, en tant que secrétaire de direction, en classes à effectif réduit et en classes d'accueil pour enfants migrants.

4.2 Les quatre enseignants de mathématiques

- Philémon : Diplômé de l'Ecole Normale, puis titulaire de brevets de maître de classe primaire supérieure I et II (BFC), il enseigne depuis dix-neuf ans. Il a, parallèlement à son enseignement, suivi des cours à la Faculté de Psychologie et des Sciences de l'Education de l'Université de Genève durant cinq ans, sans toutefois les certifier. Il occupe deux mandats : doyen du cycle I à hauteur de 50% et enseignant à 50%, avec des heures de mathématiques en 11 VSO et d'appui en 7-8⁹⁰.
- Catherine : Titulaire d'un brevet d'enseignement primaire de l'Ecole Normale, elle enseigne depuis douze ans. Elle a travaillé deux ans à l'OPTI⁹¹, et essentiellement dans les anciens degrés 7 à 9 en VSO. Catherine occupe un mandat de professeure référente pour les enseignants novices arrivant dans l'établissement. Elle enseigne à 70% en 11 VSO les mathématiques, la comptabilité, la citoyenneté et l'approche du monde professionnel.
- Paul : Titulaire d'un Master en enseignement obtenu à la HEP Vaud, il est au bénéfice de trois ans d'expérience. Avant de se former à l'enseignement, il a obtenu un CFC de cuisinier, puis un diplôme d'ingénieur. Il est maître de classe dans une 10VG où il enseigne les mathématiques au niveau 2, ainsi que la physique, les sciences, l'approche du monde professionnel et le projet interdisciplinaire, pour un plein temps.
- Fernand : Il enseigne dans un établissement différent. Titulaire d'un Master en Biologie de l'Université de Lausanne, il s'est également formé à la gestion d'entreprise et à la gestion de projet. Il a ensuite obtenu son titre d'enseignement au séminaire pédagogique pour l'enseignement secondaire (SPES). Il enseigne depuis quatorze ans, mais il a

⁹⁰ Suite à notre collaboration dans cette thèse, Philémon a été engagé comme expert aux examens oraux d'un module portant sur l'évaluation à la HEP. Il occupe toujours les mêmes fonctions au sein de son établissement.

⁹¹ L'Organisme pour le Perfectionnement scolaire, la Transition et l'Insertion professionnelle (OPTI) accueillait chaque année quelque mille élèves, âgés de 15 à 18 ans, sur 7 sites. Il visait à optimiser la transition entre l'école obligatoire et la vie professionnelle pour les élèves sans perspectives. Renommé Ecole de la Transition (EdT) dès la rentrée 2017, ses finalités sont restées plus ou moins les mêmes, mais son organisation a quelque peu changé (<https://www.vd.ch/themes/formation/apprentissage/les-ecoles/de-la-transition/>).

engrangé diverses autres expériences professionnelles : guide naturaliste, assistant à l'Université, doyen pédagogique. Il occupe un poste à temps plein, avec des heures de mathématiques et de sciences en 9VP, en 10G niveau 2, et du co-enseignement en classe d'accueil. Il est par ailleurs coach scolaire au sein de son établissement à hauteur de huit heures par semaine, pour ses collègues et pour les élèves, fonction pour laquelle il est en cours de formation⁹².

Nous remarquons que les cahiers des charges des enseignants engagés dans la recherche les font intervenir souvent dans de nombreuses classes, dans diverses disciplines et dans plusieurs niveaux, et parfois dans deux organisations du système différentes⁹³. Par ailleurs, la majorité d'entre eux exerce des mandats au sein de leur établissement.

5 Le déroulement de la recherche

Dans un premier temps, nous expliciterons le dispositif complet. Dans un second temps, nous en détaillerons chaque phase, qu'elle soit relative au volet formation ou davantage à celui de la recherche.

5.1 Un dispositif de recherche-formation dual

D'un point de vue de recherche, nous avons construit notre dispositif de façon à ce qu'il nous permette d'accéder à des données en lien avec nos questions de recherche, tout en étant en adéquation avec notre cadre méthodologique (voir partie 6). Ce dispositif s'appuie sur un dispositif de formation continue pré-existant que nous mettons en oeuvre depuis plus de dix ans. Les trois demi-journées de formation en ont donc constitué la colonne vertébrale, puisque notre but était de mener à bien une recherche collaborative⁹⁴ au coeur des pratiques réelles des enseignants. C'est dans un second temps que nous avons apparié à ce volet formation un autre volet recherche, en tentant de les articuler. Dans l'idée d'entrer dans la « boîte noire » des participants afin de mieux comprendre leurs pratiques évaluatives, nous avons décidé de mettre en oeuvre cinq autres moments, lors desquels des données qualitatives seront récoltées : deux

⁹² Un an après la prise de données, Fernand a quitté l'enseignement obligatoire pour se destiner aux classes MATAS (Module d'activités temporaires alternatives à la scolarité). A notre connaissance, il y travaille toujours.

⁹³ Au moment où se déroule cette recherche, la mise en oeuvre de la nouvelle Loi sur l'Enseignement Obligatoire (LEO) est en cours et ne touche que les degrés 9 et 10 de la scolarité obligatoire (anciens degrés 7 et 8). La dernière année, la 11^e, est encore sous le régime de l'ancienne loi scolaire, avec trois filières (VSO, VSG et VSB), alors que les deux années précédentes n'en ont plus que deux (VG et VP).

⁹⁴ Cette méthode de recherche est décrite théoriquement en détail dans la partie suivante.

entretiens semi-dirigés dont un pré et un post formation, et trois écrits exemplaires de pratiques dont un en amont, un en cours et un en fin de formation⁹⁵. Une séance appelée « zéro » a eu lieu en tout début de démarche, après les entretiens d'entrée en matière. Elle s'est avérée nécessaire pour que les participants se rencontrent et fassent connaissance, pour que les points administratifs puissent être abordés. La nature des séances, leur agencement temporel et leur contenu sont présentés dans le tableau ci-dessous.

Tableau 3: Le dispositif de recherche-formation, déroulement chronologique

⁹⁵ Le lecteur se référera également à la partie suivante pour les éléments théoriques relatifs à ces données.

Tableau 3 : Le dispositif de recherche-formation, déroulement chronologique

Temps	0	1	2	3	4	5	6	7	8
Date	10 décembre 2014 1600-1730	Négocié et mené avec tous les enseignants	Négocié et mené avec six enseignants	12 février 2015 830-1200	Négocié et mené avec six enseignants	MH : 24 mars 2015 830-1200 / FR : 23 mars 2015 830-1200	4 mai 2015 830-1200	Négocié et mené avec six enseignants	Négocié et mené avec tous les enseignants
Séances	<i>Séance préparatoire</i>	<i>Entretien individuel 1</i>	<i>Récit de pratique individuel 1</i>	<i>Formation partie 1</i> <u>Commune</u>	<i>Récit de pratique individuel 2 dans la formation avec une temporalité au choix des enseignants</i>	<i>Formation partie 2</i> <u>Différenciée MH⁹⁶ et FR⁹⁷</u>	<i>Formation partie 3</i> <u>Commune</u>	<i>Récit de pratique individuel 3</i>	<i>Entretien individuel 2</i>
Contenus	<i>Présentation de chacun.e / Présentation et négociation du dispositif / Fixation des dates / Choix des accompagnements individuels / Négociation des contenus / Questions sur l'évaluation / Points administratifs</i>	<i>Fondé sur un canevas pré-établi</i>	<i>Fondé sur un canevas pré-établi</i>	<i>Analyse d'une évaluation sommative / Modélisation / Apports théoriques / Création d'une première évaluation sommative (à envoyer au formateur avant la séance suivante)</i>	<i>Fondé sur un canevas pré-établi</i>	<i>Par discipline : Discussion des évaluations sommatives réalisées / Régulations / Création d'une seconde évaluation sommative</i>	<i>Discussion des évaluations sommatives réalisées / Régulations / Confrontation des apports inter-disciplines / Bilan en termes de formation en référence aux pratiques évaluatives</i>	<i>Fondé sur un canevas pré-établi</i>	<i>Fondé sur un canevas pré-établi</i>

⁹⁶ Mathématiques.⁹⁷ Français.

5.2 Les différents temps du dispositif de recherche-formation

Comme indiqué dans le tableau, huit temps structurent notre dispositif de recherche-formation, dont certains sont communs : entretiens semi-dirigés, récits de pratiques et séances de formation.

5.2.1 Temps 0 : une séance préparatoire

Cette séance d'une heure et demie a permis de fédérer le groupe. Elle avait également pour but de récolter d'autres questions, problèmes ou besoins des enseignants. Le plan de séance précis peut être consulté en annexe Partie 6, 1.

5.2.2 Temps 1 et 8 : des entretiens individuels semi-dirigés

Les entretiens individuels semi-dirigés, pré et post formation, se sont déroulés soit au domicile des enseignants, soit sur leur lieu de travail. Suivant un questionnement visible dans l'annexe Partie 3, 1 et ancrés méthodologiquement dans la partie 6, ils ont duré entre 65 et 85 minutes. Le premier avait essentiellement pour but de faire un état des lieux des pratiques, connaissances, croyances, questions ou autres problèmes de chacun. Le deuxième avait davantage une fonction de bilan du parcours réalisé.

5.2.3 Temps 2, 4 et 7 : des récits de pratique

Il a été proposé à tous les participants de rédiger des écrits de pratique. Six enseignants sur huit se sont portés volontaires, dont toutes les enseignantes de français⁹⁸. Autrement dit, seuls deux enseignants en mathématiques ont accepté de rédiger les textes⁹⁹. Des canevas d'écriture, présentés en annexe Partie 3, 2 et suivant une progression, ont été donnés aux enseignants. Le premier écrit avait pour intention de permettre à chacun de raconter une situation pédagogique vécue dans sa propre pratique d'évaluation sommative. Le critère de choix de cette situation devait être le fait qu'elle était particulièrement pertinente, faisait sens à leurs yeux. L'essentiel

⁹⁸ Les deux enseignants de mathématiques qui ont refusé d'entrer dans cette démarche l'ont fait pour des questions de rapport à l'écrit qu'ils ont décrit comme difficile. Nous avons respecté leur choix, de manière à rester cohérent avec notre méthode de recherche collaborative.

⁹⁹ Nous précisons ici qu'après une première recension des données, nous avons choisi de ne considérer celles ne concernant que six enseignants, deux d'entre eux n'ayant pas écrit de texte, et de surcroît, peu participé lors des séminaires.

était que chacun rende compte d'une réalité, d'un questionnement, de réflexions. Le deuxième texte avait pour finalité de mieux appréhender le processus d'appropriation en cours, lié à la formation en train d'être vécue. Nous avons demandé d'écrire un texte narratif qui montre quels impacts, positifs ou négatifs, les apports de la formation avaient sur la pratique évaluative sommative de chacun. Enfin, le troisième et dernier texte, produit à la fin de la formation, visait à dresser un bilan des apports de la formation : comment les contenus travaillés pouvaient permettre à chacun d'aborder les questions d'évaluation sommative dans sa pratique suite à la formation ? Nous avons demandé aux enseignants de formuler tous leurs textes en « Je », de manière à les rendre acteurs de leurs propos, et à privilégier une démarche d'écriture très ouverte, d'un point de vue de forme et de fond. L'ancrage méthodologique de la mise en oeuvre des récits est également explicité dans la partie 6.

5.2.4 Temps 3 : une première réunion de formation

Ce temps marque l'entrée formelle dans le processus de formation au coeur de notre recherche. Le plan détaillé de la séance et les objectifs de formation poursuivis sont présentés en annexe Partie 6, 1.2. Ce dispositif a été élaboré sur la base d'un dispositif préexistant, mais il a subi deux modifications majeures. La première est liée à l'ajustement et à la prise en compte des problèmes recensés par les participants lors des rencontres préliminaires et du temps « zéro ». Il nous a fallu en tenir compte, afin d'être proche de leurs expériences, de leurs réalités contextuelles. La seconde modification tient dans le choix des évaluations à analyser, les participants ayant explicitement demandé de travailler sur certains objets disciplinaires¹⁰⁰.

Quatre moments marquent cette première demi-journée. L'annexe Partie 6, 1.2.1 les présente en détail :

Moment 1 : présentation du PER

Le but est ici de voir comment les enseignants réagissent face au PER, de documenter les représentations et compréhensions qu'ils en ont, de partager une définition commune des termes qui y circulent, et de saisir jusqu'où il constitue pour eux un référentiel d'évaluation¹⁰¹.

¹⁰⁰ Ce qui marque de notre point de vue la cohérence avec notre méthode de recherche collaborative, décrite dans la prochaine partie.

¹⁰¹ Bien qu'il n'en soit pas un, au sens formel et théorique du terme, comme nous le développons dans notre partie conceptuelle (Partie 3).

Moment 2 : analyse d'évaluation élaborée par d'autres enseignants selon des consignes spécifiques

Cette évaluation sommative portant sur un objet de savoir choisi par les enseignants, provenant d'un autre contexte, et fournie par le chercheur-formateur. Par expérience, nous nous sommes rendu compte que cette phase mérite d'être réalisée sur une évaluation que n'ont pas produit les participants. Cette distanciation autorise un relâchement, une audace dans l'analyse, puisque personne n'est directement concerné. Le jugement, tout en restant bienveillant, peut aller en profondeur. Il est dès lors possible, d'un point de vue de formation, de voir quels objets d'analyse font sens pour les enseignants, sur quels points saillants ils s'attardent, mais aussi d'identifier les tensions, problèmes et liens qu'ils réalisent avec leurs propres situations. D'un point de vue de recherche, il rend possible la récolte d'éventuelles données portant sur les pratiques.

Moment 3 : présentation et discussion du modèle théorique proposé

Cette phase est particulièrement importante et délicate. Importante, car elle est le lieu du transfert de l'analyse à une forme de modélisation de la cohérence d'une évaluation. L'alignement curriculaire élargi, ici, permet d'objectiver les phénomènes relevés, de les faire exister. Mais cette phase est aussi délicate, car elle confronte les enseignants à un travail d'abstraction, de distanciation, et à l'appropriation d'un vocabulaire théorique auquel ils ne sont pas habitués. Vu cette complexité, et sachant que des résistances peuvent apparaître chez les enseignants dans cet exercice de conceptualisation, il est important de débattre de cette modélisation, d'en vérifier la compréhension, d'en souligner également les forces et les limites potentielles, dans une perspective compréhensive et surtout pas prescriptive.

Moment 4 : début de l'élaboration d'une évaluation sommative selon des consignes spécifiques.

Ce travail se réalise par groupes de disciplines (annexe Partie 6, 1.2.7). C'est ici que l'on peut observer comment les enseignants vont faire cohabiter leurs références, connaissances, expérience ou habitudes avec les apports théoriques fraîchement introduits. Un accompagnement du chercheur-formateur est ici essentiel, notamment au niveau de l'appropriation du modèle théorique et de ses conséquences directes sur le travail de construction d'évaluation.

La séance se clôt avec un moment de synthèse, au travers de questions de compréhension des participants et d'institutionnalisation du formateur¹⁰². Ce dernier clarifie les modalités de retour des évaluations, de façon à ce qu'il puisse les discuter en vue de la deuxième rencontre. Il est prévu que les enseignants terminent le travail hors du temps alloué au séminaire, et idéalement, soumettent leurs élèves à l'évaluation ainsi construite, avant la deuxième séance.

Précisons enfin qu'il nous a semblé pertinent de mener cette première rencontre de formation avec tout le groupe, partant de l'idée que notre modèle théorique concerne toutes les disciplines scolaires. C'est lors de la deuxième demi-journée de formation que nous avons privilégié un travail en sous-groupes de spécialistes, afin, justement, d'y approfondir l'analyse des phénomènes évaluatifs en référence aux objets disciplinaires. C'est cette rencontre que nous détaillons maintenant.

5.2.5 Temps 5 : deuxième séance de formation

Ce temps a été ajouté suite aux réaménagements réalisés sur le dispositif de départ, afin de permettre aux enseignants de présenter dans le détail leurs démarches de construction d'évaluation, et de les confronter aux remarques, questions, suggestions de leurs pairs spécialistes de la même discipline. Le fait de privilégier un tel moment nous a semblé fondamental, dans la mesure où il participe à mieux comprendre le rôle que joue la discipline dans leurs pratiques évaluatives, notamment au regard des objets et de leurs caractéristiques.

La séance commence par la présentation des objectifs de formation poursuivis (annexe Partie 6, 1.3, comme tous les autres documents utilisés à ce moment-ci). Viennent ensuite les présentations des évaluations réalisées (trois pour le français, et trois pour les mathématiques, certains enseignants ayant travaillé en duo). Il s'agit pour les concepteurs d'explicitier le cheminement parcouru dans l'élaboration des épreuves, en ayant comme fil rouge compréhensif l'alignement curriculaire élargi. Vient ensuite la confrontation de leur compréhension avec la discussion émanant du chercheur-formateur (annexe Partie 5, 9). Un point de situation en référence aux questions communes (annexe Partie 6, 1.3.3) clôt ce moment. La confrontation à ces questions, issues de la séance précédente, poursuit trois buts :

- Proposer une forme d'institutionnalisation du chercheur-formateur au travers de points de problématisation en référence au modèle théorique. Par exemple, comment

¹⁰² Cette posture particulière du formateur au sein d'une méthode de recherche collaborative est problématisée dans la partie suivante et discutée dans notre conclusion finale.

comprendre que tous les enseignants soient préoccupés par les questions de seuils dans la construction des notes ?

- Recenser les questions qui semblent résolues pour les participants et celles qui résistent ;
- Susciter un nouveau questionnement, basé sur celui préexistant. L'idée est ici de voir dans quelle direction l'appropriation des contenus et la pratique évaluative évoluent, quels points d'attention, voire de tensions, dominent.

Suite à ce travail, nous présentons la taxonomie d'Anderson et Krathwohl (2001). C'est à nos yeux un moment également très important. En effet, il va permettre de conceptualiser la notion de cohérence au sein de l'alignement curriculaire élargi, qui est centrale dans notre modèle comme nous l'avons montré dans notre cadre conceptuel (partie 3). À ce titre, cette taxonomie constitue le second apport théorique phare de la formation. Là aussi donc, nous le mettons en débat, afin de voir quelle compréhension en développent les enseignants, d'observer en quoi cet apport fait sens pour leur pratique ou comment la pratique le questionne, et voir si des résistances apparaissent.

5.2.6 Temps 6 : troisième séance de formation

Cette demi-journée de formation, la dernière du dispositif, ressemble à de multiples égards à la précédente, décrite sous le temps 5. Sa spécificité tient dans une phase de présentation et de discussion collective des évaluations réalisées en sous-groupes de discipline. Ce moment est suivi d'un temps lors duquel un enseignant de français présente son travail à l'ensemble des participants, suivi d'un enseignant de mathématiques. L'idée est notamment ici de rendre possible une mise en débat d'éléments de validation des évaluations transversaux aux disciplines, permettant de déboucher sur les questions communes relevées en début de formation : sont-elles résolues ? Quels points de tensions subsistent et pourquoi ? Quels éléments font sens, jusqu'où ? Autant de questions que l'on aborde en guise de synthèse. Ce travail est mené avec l'accompagnement du chercheur-formateur.

6 En guise de transition

Comme mentionné dans l'introduction de cette partie, nous avons construit notre dispositif de façon, d'une part, à ce que les contenus de formation puissent y circuler, et d'autre part, d'un point de vue de recherche, de manière à récolter des données méthodologiquement cohérentes. En effet, pour la formation, nous avons privilégié une méthode de travail articulée autour du

concept d'alignement curriculaire élargi, et pour la recherche, un dispositif à visée compréhensive, dans une perspective collaborative que nous présentons dans la partie suivante.

Partie 6 - Ancrage méthodologique

1 Introduction

La méthodologie se définit comme la « stratégie, le plan d'action, le processus sous-jacent aux choix et à l'application de techniques de travail spécifiques nommées méthodes. Elle fait le lien entre le choix des méthodes et les résultats attendus » (Crotty, 1998, p. 3, cité par Karsenti, 2006). C'est la raison pour laquelle nous commencerons par une partie décrivant notre méthode de recherche.

Nos choix méthodologiques prennent appui sur des constats inquiétants relevés par la recherche: depuis une vingtaine d'années, un risque de déprofessionnalisation menacerait les enseignants (Malet, 2015). Ce risque se traduirait par leur démotivation grandissante, due essentiellement à une dépossession de leur pouvoir d'action par la noosphère, les assimilant davantage à des techniciens qualifiés de l'éducation qu'à des praticiens réflexifs et responsables (Osborne, 2002 ; Rayou, 2008). En 1996, Perrenoud tirait déjà la sonnette d'alarme en affirmant que le politique, aux prises avec une profonde mutation du monde de l'école et de la formation, devait choisir un modèle de système éducatif valorisant la professionnalisation des enseignants¹⁰³. Touchant de multiples dimensions de la pratique enseignante, ce risque de déprofessionnalisation (Perrenoud, 2010) a été spécifiquement pointé par des travaux relativement aux compétences évaluatives des enseignants. Des chercheurs, comme Maroy et Cattonar (2002) ou Osborne (2006), ont montré que ces derniers se sentaient dépossédés de leurs compétences à évaluer leurs élèves lorsqu'ils étaient confrontés à de nombreuses et récurrentes modalités d'évaluation externe. Idée que soutient Jorro (2015), lorsqu'elle affirme que « des pratiques soumises à des injonctions externes, comme les évaluations internationales, engendrent un risque de déprofessionnalisation ». Lessard parle, à ce niveau, de « déqualification » (1999, p. 5), terme fort s'il en est.

Ces constats rejoignent des observations que nous réalisons depuis une dizaine d'années sur le terrain de manière informelle, lors d'actions de formation continue, et que nous détaillons dans la partie précédente ainsi que dans notre introduction. Une forme de découragement, d'épuisement, voire de résignation sembleraient gagner de nombreux enseignants lorsqu'ils

¹⁰³ En 2010, à la tête du laboratoire Innovation-Formation-Education (LIFE), il interrogeait les conditions nécessaires à la professionnalisation des enseignants et questionnait la capacité de l'Université à les former dans ce sens.

sont confrontés à trop de changements ou de prescriptions, notamment en ce qui concerne l'évaluation sommative.

Il y aurait donc urgence à mettre au centre des dispositifs de formation et de recherche des modalités offrant notamment aux enseignants la possibilité de « s'orienter vers la résolution de problèmes, d'être autonomes dans la transposition didactique et le choix des stratégies pédagogiques, capables de travailler en synergie dans le cadre des établissements et d'équipes pédagogiques, organisés pour gérer leur formation continue » (Perrenoud, 1996, p. 554). Cette perspective met entre autres au coeur du débat sur la formation le statut de la théorie dans le processus de professionnalisation. En effet, à ces fins, il semble incontournable d'outiller conceptuellement les enseignants afin qu'ils soient acteurs de leurs décisions en s'appuyant sur des connaissances et des compétences distancées du sens commun. Si cette vision de la formation est largement partagée aujourd'hui¹⁰⁴, les modes opératoires, quant à eux, divergent, et la manière de concevoir le rapport entre théorie et pratique à ce niveau soulève de multiples questions. Par exemple, la recherche a abondamment documenté la difficulté de former les enseignants en s'appuyant sur des modèles théoriques distancés de leurs préoccupations. Des travaux ont par exemple soulevé le faible impact positif que des dispositifs de formation de type « *top-down* », ou descendants (Jorro, 2013), avaient sur les pratiques réelles : « Many studies have shown that the transfer of theory to practice is meager or even non-existent » (Korthagen, 2001, p. 2). Comme Altet (2000) ou encore Rabardel (2005), nous faisons donc l'hypothèse que savoir analyser les pratiques en situation renvoie à la capacité de mobiliser des concepts issus de la recherche étant susceptibles d'être « des instruments du pouvoir d'agir des acteurs de terrain » (Vinatier & Morrissette, 2015, p. 142).

Ces éléments mettent en évidence trois dimensions clés nécessaires au développement de compétences chez les enseignants :

- La prise en considération de leurs pratiques réelles ;
- Une attention portée aux questions issues de leur réalité de classe ;
- Une distanciation d'un modèle surplombant ou le chercheur-formateur¹⁰⁵ et l'enseignant seraient au coeur d'une relation asymétrique dans laquelle le premier détiendrait l'expertise et le second serait le relais de l'expérience.

¹⁰⁴ Notre partie (5) présentant le cadre de conceptuel de l'*assessment literacy*, modèle que nous exploitons pour documenter le développement des compétences évaluatives des enseignants, expose notre position sur ce sujet.

¹⁰⁵ Nous explicitons méthodologiquement plus loin cet appariement de termes répondant à une nouvelle fonction de l'intervenant. Nous revenons également sur cette fonction dans notre conclusion finale.

Une méthode de recherche plutôt ascendante, visant l'implication des enseignants et proche d'une modalité d'accompagnement des acteurs (Charlier & Biémar, 2011) s'est alors dessinée : la recherche collaborative, qui reprend ces éléments pour en faire autant de piliers.

Cette partie explicitera donc finement cette méthode de recherche que nous avons adoptée tout en l'adaptant quelque peu à notre objet et à notre contexte pour des questions de validité méthodologique. Puis, nous aborderons les perspectives compréhensive et interprétative de notre recherche afin de mettre en évidence notre orientation épistémologique. Une fois nos quatre sources de données décrites, nous poursuivrons notre exposé en présentant nos méthodes d'analyse à l'aide des catégories conceptualisantes de Paillé et Mucchielli (2013) et de la perspective théorique de *l'assessment literacy* (Edwards, 2017 ; Willis, Adie & Klenowski, 2013)¹⁰⁶. Nous nous attarderons ensuite sur les processus de triangulation originaux mis en œuvre, afin de valider et de donner à voir le plus finement possible la complexité des résultats. Nous terminerons cette partie par une présentation des deux axes d'analyse choisis.

2 Les raisons de mener une recherche collaborative

Les méthodes de recherche, lorsque l'on souhaite investiguer les pratiques réelles, sont potentiellement multiples. Nous avons souhaité ici, dès le départ, inscrire la nôtre dans un dispositif premier de formation que nous mettons en œuvre dans les établissements depuis plus de dix ans¹⁰⁷. Ce dispositif donne une place privilégiée à la prise en compte des pratiques évaluatives réelles des enseignants et à l'expression des problèmes qu'ils y rencontrent. Il offre par ailleurs la possibilité aux enseignants de s'approprier des connaissances théoriques pour analyser, concevoir et discuter de leurs évaluations sommatives. Notre dispositif de recherche-formation articule donc des moments de formation exploités dans des visées de recherche (les séminaires) avec des phases de recherche permettant des actions de formation (les écrits et les entretiens)¹⁰⁸. Cette conjugaison a abouti à une méthode de recherche collaborative.

2.1 La genèse de la recherche collaborative

La recherche collaborative a très vraisemblablement vu le jour aux Etats-Unis et s'est d'abord

¹⁰⁶ Ses caractéristiques conceptuelles sont exposées dans la partie 4. Nous ne faisons ici qu'explicitier ses finalités méthodologiques.

¹⁰⁷ Nous en expliquons la genèse dans la partie précédente.

¹⁰⁸ Pour la description complète et exhaustive du dispositif en question, se référer également à la partie précédente. En ce qui concerne les objectifs de formation poursuivis avec les enseignants, voir l'annexe Partie 6.

développée en Amérique latine au début des années 1990 (Anadon & Savoie-Zajc, 2007 ; Morrisette & Tessaro, 2015). Sa naissance et son développement auraient bénéficié d'une montée des luttes populaires, de l'essor des mouvements sociaux ainsi que de la critique de production des connaissances en sciences humaines peu aptes au développement des pratiques professionnelles.

En Amérique du Nord, la recherche collaborative est aujourd'hui inscrite comme modèle dans les *school-university Collaborative Action Research model* (CAR) qui valorisent les communautés d'apprentissage interprofessionnelle pour aider les débutants à construire leur professionnalité en articulant questions de la pratique et apports théoriques (Mitchell, Reilly & Logue, 2009). Selon ces chercheurs, « CAR establishes a meaningful and explicit web of connection between research, theory, and practice » (p. 349). Au Canada, qui abrite bon nombre de chercheurs l'ayant développée, elle fait l'objet d'un renouveau, suite aux velléités des organismes de formation de produire des connaissances pertinentes pour la pratique, issues de résultats de recherche.

En Europe francophone, cet engouement grandissant pour la recherche collaborative se concrétise notamment dans le réseau « Recherches collaboratives sur les pratiques évaluatives (RCPE) » de l'association pour le développement des méthodologies d'évaluation en éducation (Admée), qui réunit un nombre conséquent de chercheurs ayant réalisé des travaux faisant maintenant référence, comme ceux de Lucie Mottier Lopez ou de Catherine Van Nieuwenhoven. Créé en 2008, il met en avant l'enjeu de développement professionnel de ce type de recherche, ainsi que l'activité de co-production de connaissances entre chercheurs et acteurs de terrain¹⁰⁹.

Cette modalité de recherche est par ailleurs vivement encouragée par divers organismes internationaux, tels l'ONU, l'OMS ou encore l'Europe à l'Horizon 2020, qui postulent que le changement, à certaines conditions que nous définissons plus loin, ne peut advenir que par les acteurs.

En sciences de l'éducation, il existe donc un intérêt marqué pour cette modalité de recherche qui vise conjointement le développement de connaissances pour la pratique et pour la science

¹⁰⁹ Pour célébrer ses dix ans d'existence en 2018, le réseau a organisé un symposium lors du colloque de l'Admée au Luxembourg sous le thème « dix ans de recherches menées au sein du réseau RCPE : quels rapports entre recherches collaboratives et évaluations ? », afin de dresser des bilans mais surtout des perspectives de recherches collaboratives pour le futur.

(Bednarz, 2013, 2015 ; Desgagné, 2001 ; Bourassa, Leclerc & Fournier, 2012), et qui conçoit ce processus comme devant être une co-construction entre acteurs issus de la pratique et de la recherche. De nombreux écrits ont été récemment produits, comme des *Handbooks* (e. g., Reason & Bradbury, 2008), pour donner à voir différentes conceptualisations de la recherche collaborative. L'enjeu majeur que tentent de relever la majorité de ces textes est de savoir comment la recherche collaborative permet aux savoirs théoriques de faire leurs preuves sur les terrains de l'expérience (Rey, 2014) et dialectiquement, en quoi les savoirs d'expérience peuvent venir nourrir les modèles théoriques (Mottier Lopez & Morales Villabona, 2016).

2.2 Les caractéristiques de la recherche collaborative

Commençons par une définition, empruntée à Desgagné (1997), chercheur reconnu qui a abondamment étudié et mis en oeuvre la recherche collaborative au Québec : « A sa base, nous dirons qu'une recherche collaborative suppose la contribution des praticiens enseignants à la démarche d'investigation d'un objet de recherche, démarche le plus souvent encadrée par un ou plus d'un chercheur universitaire » (p. 372). Van Nieuwenhoven et Colognesi (2015) complètent cette définition en précisant les deux finalités incontournables de la recherche collaborative : « faire émerger une production de connaissances, et viser un développement professionnel de chaque catégorie d'acteurs concernés par le processus de recherche » (p. 105). Ces éléments nous aident à identifier les traits caractéristiques de cette modalité de recherche, et ainsi de les discuter, d'un point de vue général, mais également spécifique à notre problématique de recherche. Elle autorise une schématisation de notre cadre descriptif de recherche collaborative, en nous inspirant des travaux de Desgagné, Bednarz, Lebuis, Poirier et Couture (2001). Nous reprendrons ensuite chacune de ces dimensions afin de les expliciter d'un point de vue général, puis plus précisément en référence à notre travail.

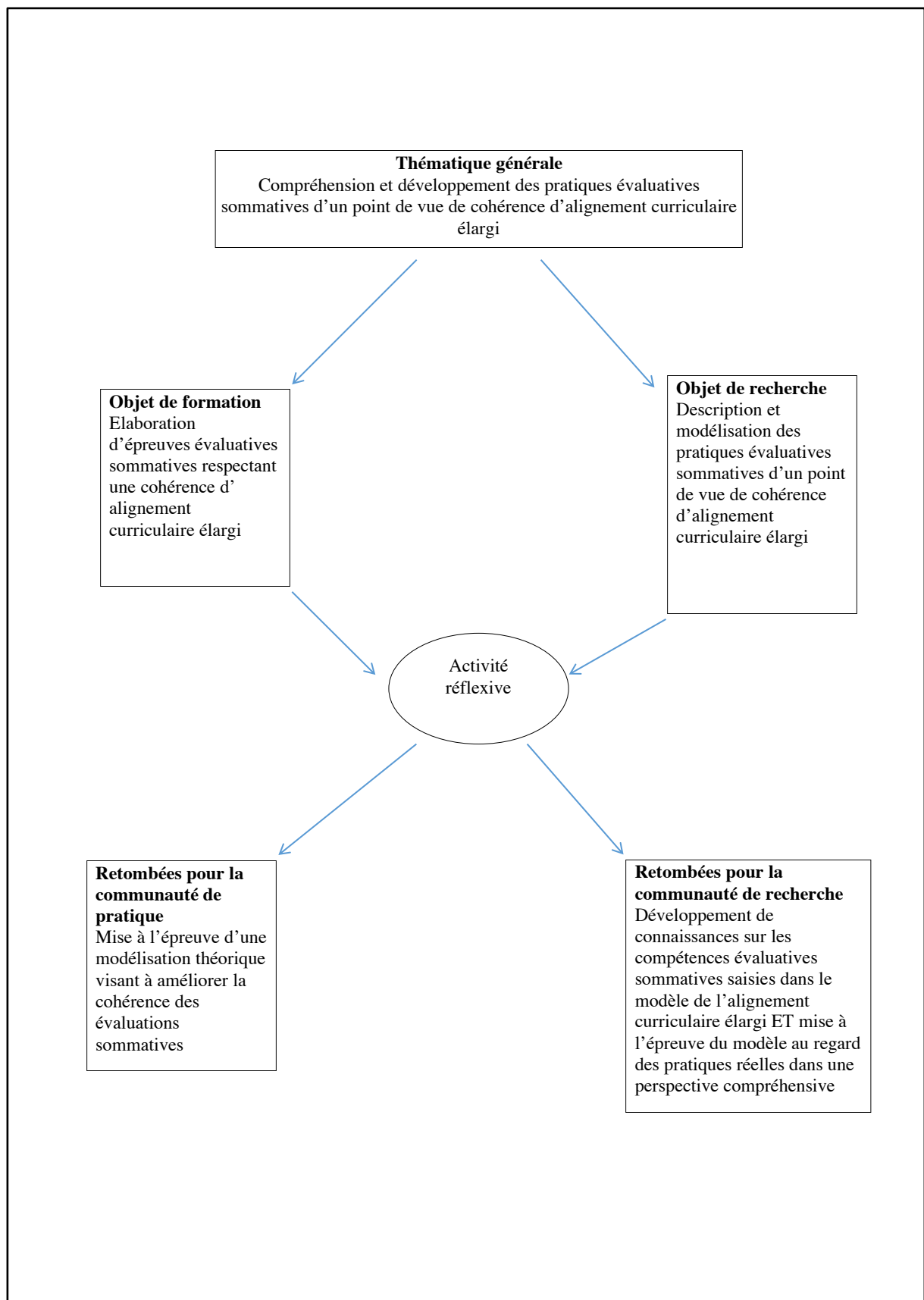


Figure 6 : La modélisation de notre recherche collaborative selon Desgagné et al. (2001)

2.2.1 La thématique générale de la recherche

C'est autour d'elle que généralement les deux parties - enseignants et chercheurs - se rencontrent pour arriver à préciser, à l'issue d'une négociation, l'objet défini dans la double dimension de recherche et de formation. Cette négociation caractérise une étape de cosituation - un espace d'intéressement, au sens de Kahn (2010) - au cœur de laquelle praticiens et chercheurs définissent un objet de la pratique à investiguer (Desgagné et al., 2001).

Cette étape est fondamentale (Nizet & Leroux, 2015). Son enjeu est double : du point de vue de la pratique, il s'agit, au travers des questionnements et réflexions partagés, de faire émerger un objet faisant sens aux yeux des enseignants pour engager un travail compréhensif et produire des connaissances sur la pratique ; du point de vue de la recherche, il est question de problématiser cet objet d'un point de vue conceptuel et méthodologique, afin de produire des savoirs scientifiques. Il s'agit donc de négocier la nature de cet objet d'investigation entre les deux parties dans une perspective de co-construction, dans notre cas : les pratiques évaluatives sommatives des enseignants et leurs différentes composantes. Fidèlement à notre méthode, son investigation se réalise *avec* les enseignants, plutôt que *sur* eux (Lieberman, 1986 ; Desgagné et al., 2001 ; Jaworski, 2003), le *avec* constituant ainsi « un élément fondateur de la recherche collaborative » (Roditi, 2013, p. 354).

Il nous est dès lors apparu fondamental de mettre en oeuvre un dispositif de recherche qui permette aux questions et aux pratiques réelles des enseignants de se confronter à un modèle théorique - puisque notre objectif est notamment de produire de la signification sur les pratiques dans une « visée herméneutique » (Roditi, 2015, p. 56) - tout en envisageant mettre également à l'épreuve de la pratique ce même modèle. En cela, nous nous distançons quelque peu du concept premier de recherche collaborative tel que Desgagné l'a modélisé (1997), puisque volontairement, nous soumettons relativement rapidement les enseignants à la contrainte de penser leurs pratiques au regard d'un apport théorique, ou, dialectiquement, de dire en quoi ce modèle est pertinent à leurs yeux pour leurs pratiques évaluatives. Cette distance, que nous assumons, nous maintient dans une méthode de recherche collaborative, puisque la co-construction de l'objet perdure, et que pratique et théorie ne cessent d'interagir pour produire du sens¹¹⁰. En effet, le modèle théorique a besoin d'être lu à l'aune de la pratique réelle pour

¹¹⁰ Rappelons que cet apport a été annoncé aux enseignants dès l'étape de co-situation et qu'ils l'ont accepté. Certains, comme Anouck ou Alizée, ont demandé dès le début à pouvoir s'appuyer sur des notions théoriques pour aller de l'avant.

épouser une épistémologie dynamique, et les enseignants ont besoin du modèle théorique proposé en formation pour développer des connaissances et des compétences distancées du sens commun.

Ce sont les raisons pour lesquelles nous considérons les pratiques non seulement au travers de ce que les enseignants font dans leur classe, c'est-à-dire de leurs connaissances mises en œuvre dans la réalité sous forme de compétences, mais également au niveau de ce qu'ils pensent et croient de leurs pratiques et de leurs contextes. À ce titre, et en référence à nos questions de recherche, nous observons les pratiques dans trois espaces privilégiés donnant lieu à une activité réflexive : des séances de formation articulant trois expressions de la pratique évaluative (analyse, conception et discussion d'épreuves), des entretiens semi-dirigés et des écrits sous forme de récits exemplaires. Nous reviendrons plus loin en détail sur ces différents éléments. Mais il est déjà intéressant de constater que c'est notre acception de la pratique, ici évaluative, qui a en partie conditionné le choix de notre méthode de recherche. Comme le dit Roditi (2015), « produire des savoirs scientifiques sur les pratiques suppose que ces dernières soient non seulement observées et explicitées, mais aussi théorisées et analysées en fonction du cadre théorique adopté » (p. 67). Il est dès lors également fondamental, dans une telle perspective, de définir le rôle du chercheur et des praticiens, ainsi que leurs modalités de collaboration au service de la production de connaissances pour la pratique et la recherche.

Notre thématique est née de constats réalisés en formation¹¹¹, raison pour laquelle elle met au centre les pratiques évaluatives sommatives réelles d'enseignants en exercice. Son inscription référée au modèle théorique de l'alignement curriculaire élargi comme modèle compréhensif¹¹² des pratiques provient également de son exploitation dans des formations continues. L'impossibilité matérielle d'investiguer plus profondément l'évolution des pratiques dans le format alors adopté (3 demi-journées de formation, avec des effectifs d'enseignants pouvant aller jusqu'à 20 personnes, toutes disciplines et degrés confondus) est à la base de cette curiosité à investiguer les questions en suspens, questions que nous avons retrouvées chez les participants à la recherche lors de la phase de négociation.

¹¹¹ Pour l'explicitation de ces constats, se référer à l'introduction de la thèse.

¹¹² Nous explicitons dans la partie conceptuelle 3 les raisons qui nous font considérer ce modèle théorique comme compréhensif, alors qu'il est décrit par certains chercheurs comme plutôt prescriptif.

2.2.2 La double dimension de recherche et de formation

Du point de vue de la formation, le fait de travailler à l'élaboration d'épreuves sommatives en lien avec l'enseignement répond à une demande explicite des enseignants exprimée lors de la phase de négociation¹¹³. Pour eux, c'est en « faisant » qu'ils verraient comment éprouver de nouvelles ou différentes manières de procéder, de trouver des réponses à leurs nombreuses questions. L'enjeu de partager une culture commune entre binômes intervenant dans une même classe ou au sein d'un établissement est également bien présent. Il y a donc eu, dès le départ, une attente forte autour d'un partage d'un langage et de références identiques, allant parfois jusqu'à des manières semblables de considérer la prescription (le Plan d'Etudes Romand, les directives départementales en matière d'évaluation), de construire des outils d'évaluation (grilles de critères, échelles) ou de négocier des consensus au niveau des exigences attendues chez les élèves (conception des seuils de suffisances en fonction des disciplines ou des objets de savoir enseignés).

Du point de vue du chercheur, le défi a été de récolter des données pour documenter et comprendre comment la cohérence des pratiques évaluatives sommatives réagit à la confrontation avec le modèle théorique de l'alignement curriculaire élargi, notamment au regard des caractéristiques disciplinaires contrastées (français et mathématiques) et des contextes des enseignants (type d'élèves, prescriptions locales, logiques d'équipes). Il était donc nécessaire que ces données documentent également les connaissances acquises par les praticiens et leur influence sur le développement hypothétique de compétences évaluatives¹¹⁴.

2.2.3 L'activité réflexive

L'aménagement de cette activité consiste en l'espace dans lequel se co-construit le savoir sur un objet lié à la pratique, issu du croisement de deux logiques, celle du chercheur-formateur et celle des enseignants. Dans ce sens, l'activité réflexive est au coeur de la recherche collaborative, en constitue le pivot (Bednarz, 2013). En d'autres termes, elle est « une zone interprétative partagée entre chercheurs et praticiens où les argumentations et les ressources des uns et des autres sont mobilisées, où les praticiens en collaboration avec des chercheurs viennent éclairer un certain objet lié à cette pratique » (Bednarz, 2015, p. 174). À ce titre, elle

¹¹³ Voir la description du dispositif dans la partie 5.

¹¹⁴ Les connaissances et les compétences étant étroitement liées, comme nous le justifions dans notre chapitre sur l'*assessment literacy*.

est également l'espace privilégié de récoltes des données.

Cette activité réflexive a traversé tout notre dispositif de recherche. En effet, dès le départ, les enseignants ont été amenés à thématiser, au travers d'un entretien et d'un écrit, les problèmes et/ou questions qu'ils souhaitaient traiter. Il s'agit bel et bien déjà d'un acte de réflexivité au sens de Schön (1983), puisque les enseignants ont été invités - mais le faisaient aussi spontanément - à questionner progressivement leurs connaissances, leurs habitudes, leurs prescriptions, leur expérience, au regard du modèle théorique présenté. Cette activité réflexive a perduré dans toutes les phases du dispositif, que ce soit dans l'analyse d'évaluations en vue de les modéliser, dans l'élaboration d'évaluations en référence à leur enseignement et à l'aune de l'alignement curriculaire élargi, dans la confrontation de leurs évaluations avec les éléments de discussion proposés par le formateur ou les pairs, ou encore dans les écrits faisant référence à la pratique évaluative¹¹⁵.

2.2.4 Les retombées pour la pratique

Les bénéfices à retirer d'une participation à une recherche collaborative devraient être préhensibles dès le départ pour les enseignants engagés. Il est dès lors important de se demander quelles retombées, même lointaines, peuvent espérer les enseignants à la fin de l'expérience collaborative à laquelle ils sont conviés (Barry & Saboya, 2015).

Cette recherche collaborative, bien que menée avec un échantillon restreint d'enseignants et sur deux disciplines au secondaire, donne à voir les conditions à partir desquelles une cohérence dans l'évaluation sommative peut être envisagée. Pour les enseignants, cela signifie qu'ils devraient être en mesure, progressivement, d'acquérir des connaissances et des compétences pour prendre des décisions évaluatives cohérentes et savoir les justifier, en fonction de leurs contraintes et de leurs contextes. Ils devraient être capables de répondre à leurs questions évaluatives ou à trouver des stratégies pour les affronter, faire face aux dilemmes qu'ils verbalisent. C'est une retombée significative qui devrait hypothétiquement leur permettre de réaliser des transferts, des généralisations. Car il n'est clairement pas question pour eux ici de figer les apports dans des produits - des « façons de faire » ou des « exemples à suivre » - mais bien davantage de les incarner dans une manière de penser la cohérence évaluative sommative du point de vue de leurs pratiques réelles, dans une perspective dynamique, contextuelle et

¹¹⁵ Pour voir où s'inscrivent ces moments de formation dans le dispositif de recherche-formation, se référer à la partie 5.

disciplinaire.

2.2.5 Les retombées pour la recherche

Les données récoltées et leur analyse devraient aider à mieux comprendre le développement des connaissances et compétences évaluatives sommatives, inscrites dans des pratiques réelles et faisant écho à de « vrais » problèmes rencontrés par les enseignants. Une telle démarche serait également à même de donner à voir comment l'appropriation d'un modèle théorique dans une visée compréhensive peut infléchir, ou non, les pratiques évaluatives sommatives, et à quelles conditions d'éventuels changements dans les pratiques sont possibles au regard des disciplines. En effet, nous faisons l'hypothèse que méthodologiquement, il serait présomptueux de considérer que des mouvements observés dans les pratiques d'un enseignant au sein du dispositif pourraient *de facto* survivre dans le terrain. Nous avons abordé cette question dans la partie conceptuelle traitant de l'*assessment literacy*, et nous y reviendrons dans notre conclusion de manuscrit.

Par ailleurs, et c'est un point fondamental, cette méthode de recherche collaborative pourrait mettre en évidence la pertinence, ou non, d'exploiter l'alignement curriculaire élargi à des fins de modélisation des pratiques évaluatives sommatives, dans des situations d'évaluation réelles, c'est-à-dire référées à des pratiques de classe contextualisées (Balach & Szymanski, 2003). L'objet s'y prête fort bien, puisque nous sommes à ce niveau en pleine adéquation avec une des caractéristiques principales de la recherche collaborative (Desgagné et al., 2001 ; Vinatier & Morrissette, 2015).

Enfin, cette manière de considérer les retombées est compréhensive et en lien direct avec notre choix de méthode de recherche, car elle permet également de considérer les limites, voire les difficultés relatives aux processus évaluatifs étudiés, notamment lorsqu'il s'agit de considérer leur mise en œuvre dans des contextes spécifiques.

2.3 Les modalités spécifiques de participation des praticiens dans une recherche collaborative

Pour l'enseignant, la recherche collaborative est une occasion de réfléchir sur son action, ou encore de s'interroger sur un aspect de sa pratique qui le préoccupe, qui lui pose problème (Desgagné, 1997). Le praticien est donc invité à faire état de son questionnement, à le partager

avec des pairs et un ou plusieurs chercheurs. Dans ce sens, la recherche collaborative se doit d'être ouverte à une complexité du réel exprimée du point de vue de la pratique. Cette ouverture, comme le disent Morrisette et Desgagné (2009), « est au coeur d'une tension entre un souci de rigueur pour toute démarche méthodologique [...] et la nécessité de maintenir une certaine souplesse pour mieux sentir le terrain » (p. 139). Car il s'agit ici bien davantage de saisir la pratique dans toutes ses dimensions pour lui donner un statut de production de connaissances, que de l'exploiter à des fins d'illustration d'une théorie surplombante. En cela, collaborer revient également à reconnaître aux praticiens un champ de compétences spécifique à mettre au service de l'objet de recherche (Desgagné & Larouche, 2010). Leur développement de compétences dépend dès lors également en partie de connaissances dont ils sont porteurs. On observe encore une fois ici que la conception de l'enseignant participant à de telles recherches s'approche de celle du praticien réflexif, envisagé comme un partenaire averti qui contribue, avec le chercheur, et au coeur d'une réflexion conjointe, au développement de la pratique (Desgagné et al., 2001). Il s'agit donc de reconnaître un savoir de la pratique, contextualisé et personnalisé¹¹⁶. Ce qui suppose, de manière dialectique, une posture également spécifique du chercheur.

2.4 Un rôle non moins spécifique pour le chercheur

La recherche collaborative est reconnue comme étant très exigeante pour le chercheur (Bednarz, 2015). En effet, il doit être habité, tout au long du processus de recherche qu'il mène avec les enseignants, par un critère de double vraisemblance du dispositif (Bednarz, 2013 ; Desgagné, 2007 ; Dubet, 1994), c'est-à-dire un souci de rendre compte conjointement des préoccupations des praticiens et des siennes, participant ensemble à la construction de connaissances pour la pratique et la recherche. Bednarz (2015) décline ce critère en fonction des étapes du processus de recherche : « double pertinence sociale lorsque se co-construit le projet, double rigueur méthodologique dans la co-activité autour des pratiques, [...] et double fécondité des résultats » (p. 181). Ce sont à notre sens les travaux de Desgagné (1997, 2001, 2007) qui ont caractérisé le mieux ces trois phases principales d'une recherche collaborative :

- Une phase de co-situation, où, comme nous le montrons plus haut, il s'agit de faire en sorte que la situation travaillée, ou que la problématique choisie, soit pertinente d'un

¹¹⁶ Notons que ce processus de reconnaissance nourrit la co-construction.

point de vue de préoccupation de recherche, mais également du point de vue de la pratique en répondant à un questionnement professionnel ancré dans une réalité vécue ;

- Une phase de co-opération où se vivent les interactions entre acteurs de la recherche, et où le chercheur favorise un espace de collecte de données en rapport avec sa méthodologie, tout en laissant ouvert un espace de questionnement de la pratique ;
- Une phase de co-production, qui rejoint la nécessité de penser une double fécondité des résultats au cœur de laquelle le chercheur reste attentif à distinguer les connaissances produites à la fois sur le plan professionnel (pour la pratique), mais aussi sur le plan académique (pour la recherche), en lien avec des questions du terrain. Le chercheur est donc amené à interpréter la parole et les actes des enseignants engagés au moment de la recherche, et dans l'analyse des données. Dans cette perspective, il participe à la construction du sens.

À l'image de Desgagné et al. (2001), nous postulons que c'est ce critère de double vraisemblance qui semble le mieux éclairer le rapport particulier entre recherche et formation présent dans notre dispositif de recherche. En effet, dans la mesure où le chercheur est confronté à la nécessité de guider les praticiens dans leur questionnement et démarches d'appropriation, il intervient au-delà de son rôle d'observateur. Dans ce sens, il investit ici également et de manière articulée, un rôle de formateur qui « encadre la démarche de recherche des enseignants, prise au sens informel du terme » (Desgagné et al., 2001, p. 39). Ce double investissement de postures considérées parfois comme peu compatibles (Bednarz, 2015) peut selon nous toutefois se réaliser aux conditions suivantes : garder constamment en vue l'objet de la recherche, éviter d'investir une posture d'expert surplombante, et rester fidèle au processus d'investigation requis par la recherche. Nous parlons alors d'une posture d'accompagnateur.

2.5 Qu'en est-il du chercheur-formateur ?

La posture adoptée par l'intervenant dans notre recherche est celle d'un chercheur qui s'implique dans le projet pédagogique qu'il accompagne (Cuche, Canivet & Donnay, 2012) : une posture, en réalité, triple :

- Il est chercheur d'une part, puisqu'il poursuit l'investigation d'un objet de recherche et a pour but de produire des connaissances scientifiques en référence à des questions précises co-construites avec des enseignants ;

- Il est formateur d'autre part, car il mène une action de formation auprès des participants visant à les faire progresser dans leurs pratiques. Sa connaissance du contexte lui permet de mieux identifier les questions et problèmes des enseignants (Tessaro, Jeannet, Favre Marmet & Pamm Wakley, 2010). Par ailleurs, il assume qu'il dispense des connaissances, notamment au travers de l'apport d'un modèle théorique, et fait l'hypothèse que la rencontre entre les questions des praticiens et le modèle permet de nouvelles acquisitions, enclanche des mouvements dans les pratiques ;
- Il est accompagnateur d'équipes d'enseignants mobilisés ici par l'expérimentation de pratiques évaluatives référencées à un modèle théorique, puisqu'il aide ces derniers à cheminer dans leur parcours individuel et collectif au regard de leurs questions, contextes, mais aussi à gérer les tensions, craintes, voire découragements¹¹⁷ auxquels ils pourraient être confrontés.

L'implication du chercheur-formateur¹¹⁸ traduit donc ici une autre épistémologie de la recherche collaborative, puisque – à l'instar des enseignants – il fonde son action sur trois axes d'intervention. Il peut dès lors à nos yeux être considéré comme un accompagnateur, c'est-à-dire un partenaire créateur de connaissances, en assumant le fait que ces connaissances sont orientées par les questions et problèmes des enseignants, mais aussi par l'ancrage épistémologique du modèle proposé. En effet, dans notre dispositif, un des rôles de l'accompagnateur est de proposer un ensemble de réflexions possiblement favorisées par l'appropriation d'un modèle théorique en référence au questionnement de la pratique. Il est donc amené non seulement à expliciter les caractéristiques conceptuelles du modèle en question, à les mettre à l'épreuve de la pratique, mais également à aider à décrire les pratiques évaluatives, à questionner le « pour quoi » et le « comment » d'un point de vue spontané puis progressivement en référence au modèle (Boucenna & Charlier, 2012). Une posture d'expert de contenu est donc totalement assumée et assumable par l'accompagnateur, puisqu'elle est au service de la compréhension des pratiques par les enseignants, et non un moyen d'asseoir une forme de relation asymétrique entre le chercheur et les participants. Comme le dit Charlier (2012) : « La pratique n'est pas spontanément enseignante. Pour être formatrice, elle doit être théorisée » (p. 130). Et c'est bien le recours à un modèle théorique qui infléchit la posture et les

¹¹⁷ Nous pensons notamment à la crainte exprimée par certains participants suite au sentiment de faire « faux », ou au découragement ressenti suite à la complexité du travail à réaliser.

¹¹⁸ Au vu de ce qui précède, nous opterons tout au long de ce travail pour cette terminologie, puisque nous réalisons une recherche dans laquelle une formation est mise en œuvre, et non l'inverse.

rôles investis par l'accompagnateur. Nous retrouvons ici un des schèmes invariants et fondamentaux des opérations qui caractérisent le métier de chercheur selon Saada-Robert et Leutenegger (2002) : la mise en place d'un dispositif de recherche incluant un dispositif de formation.

2.6 Une perspective de recherche fondamentalement compréhensive et interprétative

La visée prioritaire de la recherche collaborative est une meilleure compréhension des objets et des phénomènes sur lesquels travaillent praticiens et chercheurs. À ce titre, elle nécessite un croisement entre des compréhensions différentes (Bednarz, 2015). Notre dispositif, malgré ses particularités, induit tout de même une posture de recherche fondamentalement compréhensive. En effet, l'approche compréhensive rend compte d'un positionnement intellectuel qui postule que « les faits humains ou sociaux sont des faits porteurs de significations véhiculées par des acteurs (hommes, groupes, institutions, etc.), parties prenantes d'une situation humaine » (Paillé & Mucchielli, 2012, p. 40). Pour comprendre le sens des actions humaines, il devient dès lors important de comprendre le contexte¹¹⁹ dans lequel elles s'inscrivent, chercher le sens plutôt que l'explication. D'un point de vue méthodologique, l'étude du fait humain se base sur un fait déjà construit et donc existant dans une communauté, qui appelle des outils compréhensifs de type qualitatifs. Dans ce sens, dès lors que l'on vise à proposer des interprétations possibles des phénomènes observés plutôt que de les réduire à des faits objectivables, la posture devient également interprétative (Gohier, 2004). La recherche interprétative appelle donc également des outils qualitatifs (Guba & Lincoln, 1989).

3 L'univers interprétatif

Un enjeu méthodologique de notre recherche réside dans la validité de l'articulation entre notre modèle théorique et l'analyse qualitative de nos données issues des pratiques évaluatives observées, écrites ou verbalisées. Nous avons tenté, à ce niveau, de traduire une interprétation des phénomènes observés en croisant, dans un double mouvement décrit plus haut, un modèle théorique avec le monde du chercheur. En effet, tout chercheur est habité par une histoire personnelle et disciplinaire, par des expériences de recherche et de formation, par des angles d'approches qui, ensemble, conditionnent et orientent son « univers interprétatif » (Paillé &

¹¹⁹ Cette prégnance du contexte est également une des caractéristiques du modèle de l'*assessment literacy*, exploité pour l'analyse de nos données et dont nous présentons les caractéristiques conceptuelles dans la partie 4, ce qui est à nos yeux un gage de validité méthodologique.

Mucchielli, 2012), défini comme « la ligne d'orientation préalable et provisoire rendant possible tout notre expérience » (Gadamer, 1996). Parallèlement, notre modèle théorique vise à formaliser cet univers interprétatif pour lui donner une portée allant au-delà des objets étudiés, pour idéalement, aboutir à une généralisation de leur objectivation. De ce fait, les postures compréhensives et interprétatives « jouent ensemble » pour un gain de validité méthodologique, en s'appuyant notamment sur des données permettant leur interprétation et leur compréhension à l'aide d'outils qualitatifs.

4 Quatre sources de données

Nous avons procédé au recueil de quatre types de données : des écrits de pratique « exemplaires », des entretiens dits compréhensifs (Kaufmann, 2011), des échanges pendant les actions de formation, et des évaluations sommatives réalisées par les enseignants ainsi que leurs discussions par le formateur. Ces données se sont imposées à nous pour deux raisons.

La première tient au fait que notre méthodologie s'inscrit, comme dit plus haut, dans une démarche compréhensive et interprétative. Il s'agissait donc d'envisager des données pour à la fois « entrer » le plus possible dans les pratiques évaluatives réelles, mais aussi pour obtenir des traces de l'évolution des discours des enseignants sur leurs pratiques saisies dans notre modèle théorique. Pour cela, les entretiens ont ouvert et clos le dispositif. Les trois écrits l'ont jalonné, puisque les enseignants les ont produits avant, pendant et après la formation. Les discussions menées lors des actions de formation se sont déroulées lors des trois séminaires. Les évaluations sommatives, quant à elles, ont constitué le cœur du dispositif pour les enseignants, ont rythmé les actions de formation. À ce titre, elles donnent à voir une expression de la pratique réelle d'évaluation, puisqu'elles ont été produites en référence à l'enseignement dispensé dans les classes où elles ont été mises en œuvre, et fidèlement à leur contexte. Reprenons ces données en détail.

4.1 Des écrits pour accéder à une forme de prise de recul

Pour asseoir notre démarche compréhensive et interprétative, nous avons proposé aux enseignants de produire des récits de pratiques évaluatives. Les écrits ont été recueillis tels quels, de manière authentique, dans leur première version. Les enseignants les ont produits fidèlement à un canevas d'écriture volontairement ouvert (voir annexe Partie 3, 2).

Le récit fait l'objet d'une attention particulière dans l'étude des pratiques depuis la fin des années 1990, et semble être aujourd'hui en plein essor. Quelle que soit sa forme (portfolio, journal, texte réflexif), il poursuit une finalité de professionnalisation (Crinon & Guigue, 2006 ; Marcel 2014). Différents travaux témoignent de cet intérêt à donner la parole aux enseignants quand il s'agit de rendre compte de toute la complexité inhérente à l'exercice de la pratique professionnelle, du point de vue de ceux qui l'assument au quotidien (e. g., Cifali & Myftiu, 2006 ; Desbiens, Borges & Spallanzani, 2012 ; Desgagné, Gervais & Larouche, 2001 ; Gervais & Desgagné, 2003 ; Pasquini, 2013, 2016).

Dans un précédent ouvrage (Pasquini, 2013), nous avons souligné, en nous inspirant des travaux de Desgagné (2005), l'intérêt de considérer le récit de pratique comme porteur de connaissances expérientielles sur la profession enseignante. Nous avons également montré qu'un récit de pratique résonne avec une situation questionnante que rencontre l'enseignant, et que sa mise en mots en permet une problématisation. Dans ce sens, l'enjeu de mettre des enseignants dans une posture d'écriture de leurs expériences réside dans leur capacité à rendre compte de ce qu'ils font, disent, pensent, questionnent, relativement à leurs pratiques, ici, évaluatives. Il s'agit de leur donner l'occasion de prendre une forme de recul souvent délicate avec l'entretien afin de penser la complexité qu'ils affrontent dans toutes ses dimensions, pour pouvoir verbaliser les dilemmes, paradoxes ou certitudes qui émergent chez eux.

4.1.1 Le récit exemplaire

L'approche théorique choisie ici est celle du récit exemplaire. Pour Desgagné (2005), un récit « exemplaire » ne montre pas des exemples à suivre, mais représente des exemples d'une pratique. L'exemplarité dont il est question renvoie à la posture d'un enseignant engagé dans sa pratique, dans laquelle il a puisé des événements considérés comme significatifs. Dans ses nombreux travaux, l'auteur adopte une méthodologie d'études de cas qui s'articule autour de la résolution de problèmes exprimés par les praticiens, autour de questions de gestion de classe. Dans le cadre de notre recherche, et en nous inspirant de nos précédents travaux (Pasquini, 2013), nous souhaitons élargir cette acception. Nous postulons que le récit peut aussi être exemplaire de questions, de dilemmes, de tensions que rencontre l'enseignant dans ses pratiques et que les phénomènes qu'il interroge peuvent s'ancrer dans des dimensions didactiques ou pédagogiques. Bien que la production de ces textes ait obéi à un canevas de manière à rester proche de notre problématique, nous avons encouragé les enseignants à entrer dans une démarche personnelle et à ne pas vouloir coller à un attendu pré-défini. Comme le dit

Jorro (2002) : « Ecrire en formation représente un défi pour le formé qui passe par toutes les étapes émotives et affectives : de l'inconfort à la réassurance, de l'envie d'abandonner à l'euphorie » (p. 16). Il s'est donc agi de ne pas « brider » l'élan d'écriture.

Cette manière de procéder nous a fourni des textes très différents, tous en rapport avec les pratiques évaluatives. Certains enseignants ont en effet décrit comment ils élaboraient une épreuve, avec les questions et soucis auxquels ils devaient faire face. D'autres ont privilégié des réflexions plus « méta » en formulant des interrogations touchant plusieurs niveaux de la pratique évaluative, mais aussi des dimensions de culture de l'évaluation, de posture d'évaluateur ou encore de rapport au contexte scolaire. D'autres encore ont alterné ou conjugué les deux, dans des formes et densités très diverses. Tous ces textes sont disponibles dans l'annexe Partie 5, 5.

4.1.2 Le récit exemplaire au service de la compréhension des pratiques

Quels que soient les objets traités dans les textes, ils ont tous été produits de manière à être fidèles à une réalité vécue, ou du moins à son expression. Les objets de l'écriture entretiennent donc tous un rapport étroit avec la pratique évaluative sommative, en donnant des explicitations diverses. Dans cette perspective, les contenus des textes peuvent être lus à la lumière de catégories conceptualisantes¹²⁰, comme autant d'unités de sens identifiant des « traces des processus de réflexion » (Crinon & Guigue, 2006, p. 153) en cours. Chaque catégorie est en rapport étroit avec notre problématique et surtout avec notre modèle théorique. Ce rapport permet de mieux comprendre les phénomènes narrés, pour pouvoir ensuite les mettre en perspective avec les autres données¹²¹. La démarche méthodologique est donc ici aussi fondamentalement compréhensive puisqu'elle vise à dégager le sens de la pratique écrite, et particulièrement dynamique, car elle donne à voir le processus de construction de ce sens (Buisse & Vanhulle, 2009). Ceci dit, si le récit de pratique développe la réflexivité (Francis, 2011), il ne peut à lui seul la garantir (Crinon & Guigue, 2006 ; Guichon, 2007), raison pour laquelle nous avons procédé conjointement à des entretiens.

¹²⁰ Il nous semble important d'introduire cette dimension analytique de notre méthodologie ici, même si nous la définissons plus loin.

¹²¹ Nous parlerons à ce niveau de triangulation de méthodes, élément que nous détaillons également plus bas.

4.2 Une deuxième source de données : les entretiens

Historiquement, en sciences de l'éducation, l'entretien constitue un instrument privilégié en recherche qualitative. Plus que tout autre dispositif, il « permet de saisir, au travers de l'interaction entre un chercheur et un sujet, le point de vue des individus, leur compréhension d'une expérience particulière, leur vision du monde, en vue de les rendre explicites, de les comprendre en profondeur » (Baribeau & Royer, 2012, p. 26). Suivant Boutin (2006) ou Mucchielli (2009), nous le définirons comme une méthode de collecte d'informations qui se situe dans des interactions multiples entre un intervieweur et un interviewé en vue de partager un savoir et de dégager une compréhension d'un phénomène (Savoie-Zajc, 2009, 2016). Nous observons déjà ici la perspective compréhensive que recouvre l'entretien, dont le recours peut s'argumenter triplement (Poupart, 1997) :

- Un argument épistémologique, qui explore la perspective de l'acteur. Cette investigation en profondeur poursuit des visées de compréhension et d'explication. Elle peut être atteinte lors d'entretiens répétés et approfondis, comme dans le cadre de notre recherche ;
- Un argument éthique, qui accède à la compréhension fine des dilemmes et des enjeux que l'acteur affronte. Là, les précautions éthiques doivent être posées et explicitées, de manière à permettre au sujet de s'exprimer le plus librement possible, sans craintes ;
- Un argument méthodologique, qui donne accès à son expérience, par exemple au travers d'entretiens semi-directifs.

Dans cette perspective, plusieurs types spécifiques d'entretiens existent, dont l'entretien qualitatif serait le générique. Dans le cadre de notre recherche, et pour garder une validité méthodologique, nous avons décidé d'exploiter l'entretien compréhensif.

4.2.1 Des entretiens compréhensifs

Chaque entretien a été enregistré, puis a fait l'objet d'une transcription intégrale (annexes Partie 5 ; 1, 3, 5, 7). Nous avons opté pour un type d'entretien permettant de poursuivre nos intentions premières de recherches. Pour cela, il nous a paru essentiel d'adopter une posture d'investigation semi-directive, où, fidèle à un canevas d'entretien assez sommaire (annexe Partie 3, 1), nous avons limité nos interventions à des relances afin de susciter la parole des enseignants ou de recentrer leur propos vis-à-vis de la consigne de départ. Bien que

constamment guidés par nos questions de recherche, nous avons tenté de mobiliser les enseignants à partir de leur parole, dans un souci d'objectivité. Cette posture semi-directive est un compromis entre le « laisser-aller » de l'entretien non-directif et le dirigisme de l'entretien directif (Fugier, 2010). Il garantit dès lors une forme d'entretien relativement standardisée qui autorise une comparaison des discours recueillis, ce qui est un gage de validité.

L'entretien compréhensif prolonge l'entretien semi-directif dans la mesure où l'enseignant, ici, n'est pas considéré comme un « enquêté », mais plutôt comme un informateur, qui va possiblement faire part de *ses* raisons justifiant *ses* représentations - ce qui renvoie à la rationalité axiologique de l'acteur et à ses catégories de pensées, à partir desquelles il produit, justifie, analyse ses opinions (Kaufmann, 2011). La grille d'entretien de l'intervieweur se réfère dès lors à ses questions de recherche, tout en restant très ouvert à l'inattendu, à l'indicible, de manière à donner à l'entretien compréhensif une dimension personnelle que n'ont pas les autres types d'entretiens. À ce titre, le chercheur se doit de réaliser des allers et retours permanents « entre compréhension, écoute attentive, et prise de distance, analyse critique » (Kaufmann, 2011, p. 22). En effet, pour l'auteur, l'entretien est compréhensif, parce que son caractère familier, la dynamique de confiance existant entre les protagonistes, autorise l'intervieweur à « descendre » en profondeur. Au contraire de l'entretien traditionnel, ce type d'entretien permet d'accéder, par fragments, à une « vérité » construite par les personnes, au-delà des masques sociaux.

L'objectif principal de ce type d'entretien et de la méthode en général est de produire une théorie. Une théorie née de l'articulation entre les données et les hypothèses, elles-mêmes enracinées dans les faits vécus par les enseignants. Ce qui correspond bien à notre méthodologie de recherche, dans la mesure où c'est le questionnement de départ des enseignants qui a en partie infléchi la nature de notre dispositif.

4.2.2 De la manière de produire de la théorie

L'entretien compréhensif inverse les phases de construction classique de l'objet « élaboration d'une hypothèse fondée elle-même sur une théorie existante - procédure de vérification - rectification de l'hypothèse ». C'est le terrain qui devient le point de départ de la problématisation et qui donne des éléments de compréhension des logiques des acteurs. Dans cette perspective, ces derniers sont porteurs d'un savoir commun à forte valeur ajoutée, et l'enjeu de l'intervieweur est d'y avoir accès. D'un point de vue épistémologique, l'objectivation

se construit de manière progressive, donnant à voir l'objet de l'enquête d'une manière de plus en plus éloignée du sens commun. La problématisation de l'objet de recherche inscrit dans un tel type d'entretien est dès lors dépendante de ce foisonnement d'idées, de connaissances, de valeurs, de représentations que l'interviewé met à disposition du chercheur. En suivant toujours Kaufmann (2012), nous avons abordé les enseignants dans les deux phases des entretiens (pré et post-formation) avec l'idée que leurs pratiques évaluatives sommatives étaient en questionnement, que l'apport d'un modèle théorique pourrait leur faire réaliser des mouvements et les faire évoluer vers davantage de cohérence, mais aussi que leurs connaissances et le développement possible de leurs compétences évaluatives seraient en mesure de questionner le modèle.

4.3 Notre troisième source de données : des présentations et discussions d'évaluations sommatives lors des actions de formation

Les séances de formation 2 et 3 ont vu apparaître des moments-clés que nous avons enregistrés, puis partiellement transcrits, à savoir :

- Aux séminaires 2, différenciés français et mathématiques : la présentation des questions soulevées par les enseignants comme point de réflexion, les présentations des premières évaluations réalisées en sous-groupes avec les débats qu'elles ont suscités, et la présentation de la taxonomie d'Anderson & Krathwohl (2001) ;
- Au séminaire 3 avec tous les enseignants : les partages de pratiques en sous-groupes de disciplines au travers des présentations des secondes évaluations.

Ces données nous sont apparues comme fondamentales pour donner à voir l'évolution des connaissances et surtout leur mise en pratique sous forme de compétences des enseignants. En effet, elles participent à considérer que les enseignants, ici, sont amenés à passer d'un discours *sur* la pratique à un discours *de* la pratique, puisqu'ils mettent en débat des épreuves réalisées pour leurs élèves et les confrontent à la compréhension qu'en ont leurs pairs.

Concernant les présentations, chaque enseignant, ou duo d'enseignants, a été amené à expliquer les deux évaluations réalisées à ses collègues et au chercheur-formateur. L'idée était d'obtenir des données permettant de prendre connaissance des processus de construction, voire de régulation de ces épreuves. Relativement aux présentations, chaque évaluation a été envoyée au chercheur-formateur avant la rencontre de manière à ce qu'il puisse en prendre connaissance en amont, en vue d'y identifier les éléments de discussion pertinents à porter au sein du groupe

(voir annexe Partie 5, 9). Ainsi, chaque évaluation, comme donnée, a induit la production directe d'une donnée associée : les évaluations discutées. Construites ainsi, elles ont constitué un point d'appui clé pour confronter et mettre en discussion les pratiques établies ou en construction.

4.4 La quatrième source de données : des évaluations sommatives référées aux contextes de classe

Nous avons décrit dans le détail les moments auxquels les enseignants ont été amenés à produire, à présenter, puis à discuter deux évaluations sommatives au cœur de notre dispositif dans la partie précédente. Nous avons également dit en quoi il nous a semblé pertinent de les laisser choisir leur manière de travailler à ces réalisations, de façon à rester en phase avec leur enseignement, leurs contraintes, leurs contextes. Nous présentons ici les finalités méthodologiques poursuivies au travers de la production de ces données.

4.4.1 La production d'évaluations

Cette phase nous a permis de poursuivre deux buts :

- Accéder à une forme d'expression des pratiques évaluatives réelles saisies en contexte ;
- Obtenir des supports étayant la production d'autres données, comme les entretiens ou les écrits.

D'un point de vue méthodologique, il nous a semblé fondamental de nous référer, à un moment donné, à une forme de données explicitant concrètement les connaissances en jeu lorsque les enseignants « passaient à l'action ». Certes, on pourra nous questionner sur le fait de n'avoir travaillé que sur des évaluations sommatives « vides », sans traces d'élèves, ou d'avoir sciemment mis de côté l'analyse des processus d'enseignement et d'apprentissage les ayant précédées. Ces remarques, dans une certaine mesure, nous semblent justifiées, et constituent vraisemblablement une limite de notre recherche sur laquelle nous reviendrons dans notre partie conclusive. Cela étant, en passant par une analyse « triangulée » des données que nous présentons plus loin, nous avons pu, à notre sens, contrôler certains biais liés à ces limites. Il nous a semblé par ailleurs pertinent de voir ces productions d'évaluations comme autant de sources potentielles de réflexions exemplifiant la pratique évaluative, ces dernières étant hypothétiquement représentatives de pratiques d'enseignement. C'est la raison pour laquelle nous avons laissé la possibilité aux enseignants de s'y référer spontanément dans les entretiens,

voire dans les écrits, ce que certains d'entre eux ont fait. Leurs propos sont devenus alors plus ancrés, situés, permettant une mise en perspective allant dans le sens d'une triangulation des données.

5 Une démarche d'analyse dans un rapport dialectique entre déductif et inductif

Notre méthodologie a suivi d'abord un mouvement inductif, mettant au centre les questions et problèmes des praticiens ainsi que l'expression de leurs pratiques et leurs dimensions réelles. Cette « entrée » va dans le sens des travaux d'Anadon et Guillemette (2007) pour qui la recherche qualitative valorise « l'exploration inductive » (p. 30), la connaissance étant produite à partir de données et non soumise à un modèle théorique.

Mais dans notre cas, c'est justement un modèle théorique qui est entré en confrontation avec les questions et problèmes évoqués par les enseignants, d'abord concrètement dans le dispositif de formation, puis dans la démarche d'analyse des données. C'est alors un mouvement déductif qui a pris le relais puisque c'est à l'aune de ce modèle que les données ont été interprétées, dans une perspective compréhensive. Cette phase, pour Guillemette (2006), constitue une limite de l'approche purement inductive. Mais elle est fondamentale, car il n'est pas possible d'obtenir des résultats de recherche purement construits a posteriori. Le recours à un modèle est donc bien nécessaire en recherche qualitative : « En somme, toute approche généralement inductive implique nécessairement des moments de déduction sans perdre son caractère essentiellement inductif, celui-ci provenant de l'orientation fondamentale qui consiste à étudier les phénomènes à partir de l'expérience propre des acteurs » (Anadon & Guillemette, 2007, p. 33).

Nous observons donc une alternance entre les deux mouvements, entrant dans chaque phase de la recherche dans un rapport dialectique. De fait, notre démarche méthodologique ne peut se résumer à une opération inductive, celle-ci étant insuffisante pour dégager des éléments de compréhension et/ou d'interprétation des phénomènes observés. Mais elle n'aurait pas non plus pu se réduire à une opération déductive, tout modèle théorique devant être éprouvé au regard de données reflétant des éléments de la pratique, dans une perspective compréhensive. Dans ce sens, notre approche a mis en lien le « pouvoir explicatif » du modèle théorique avec des données construites en référence à une problématique réelle partagée par les enseignants. D'un point de vue méthodologique, c'est à nos yeux un des critères de validité de notre démarche, dans le sens où elle cible, dans une visée interprétative, le « comment », mais aussi le « pourquoi ». Comme le disent Saada-Robert et Leutenegger (2002) : « l'explication ne

suppose pas une démarche exclusivement déductive pas plus que la compréhension ne suppose une démarche exclusivement inductive » (p. 15). L'idée est plutôt ici d'ouvrir des possibles sur une potentielle reconfiguration de notre modèle théorique, tout comme d'être flexible sur les phénomènes qu'éclaireraient nos données. L'un pouvant dès lors influencer l'autre, et vice versa.

Cela étant, nous pourrions nous demander dans quelle mesure nous n'avons pas épousé une démarche abductive. L'abduction, au sens de méthode, est réalisée par une comparaison continue entre les données collectées et les construits théoriques en constante évolution (Anadon & Guillemette, 2007). Toutefois, bien que l'induction, la déduction et l'abduction aient des points communs, elles n'ont pas la même temporalité, et surtout, ne portent pas sur les mêmes objets (Dechamboux & Mottier Lopez, 2015). Dans ce sens, nous n'avons jamais eu pour but de faire évoluer notre modèle théorique, même si nous avons souhaité le soumettre aux pratiques évaluatives réelles afin de l'éprouver. Cela ne signifie en rien qu'aucun processus abductif n'a eu lieu, mais plutôt qu'il n'a pas été objectivé au point de devenir un outil de recherche. Nous resterons donc fidèle à conceptualiser le double mouvement inductif-déductif, en continuant notre propos en présentant quelques outils méthodologiques.

5.1 L'analyse à l'aide des catégories conceptualisantes

Fidèle à ce double mouvement, et afin de dégager des éléments de sens qualitatifs de nos données, nous avons opté, dans un premier temps, pour un type d'analyse à l'aide des catégories conceptualisantes (Paillé & Mucchielli, 2012 ; Raab, 2015). Ce type d'analyse nous a semblé pertinent pour trois raisons essentielles : il privilégie la dimension qualitative de la donnée ; il peut s'articuler avec une théorisation pré-existante ; son recours est valide pour toutes les données de notre recherche.

L'analyse à l'aide des catégories conceptualisantes a lieu après un premier examen empirique des données. Son objectif est de « qualifier les expériences, les interactions et les logiques selon une perspective théorisante » (Paillé & Mucchielli, 2012, p. 315). Le principe de cette méthode d'analyse réside dans une tentative de théorisation des données, afin, par exemple, de relever les processus de pensées et d'action en jeu, de qualifier les dilemmes ou les paradoxes, de cerner les logiques présentes ou les résistances dans les pratiques. Ce travail de conceptualisation et de théorisation se déroule dans le temps présent de l'analyse, à l'aide d'un outil qualitatif spécifique : la catégorie.

5.1.1 La catégorie conceptualisante

Paillé et Mucchielli (2012) définissent la catégorie comme « une production textuelle se présentant sous la forme d’une brève expression et permettant de dénommer un phénomène perceptible à travers une lecture conceptuelle d’un matériau de recherche » (p. 316). Cette méthode d’analyse nous semble particulièrement valide dans notre dispositif. Partant du postulat que toute catégorie désigne un phénomène, nous pouvons considérer les concepts de notre modèle théorique comme étant proches de catégories. Par exemple, le concept de tâche évaluative joue non seulement un rôle structurant dans la compréhension des phénomènes évaluatifs par le praticien, mais il demande au chercheur d’incarner l’attribution de signification des phénomènes relatifs au concept pour comprendre les pratiques. Aussi, dans une perspective inductive-déductive au cœur de laquelle nous avons mis en dialogue les concepts de notre modèle théorique avec nos données perpétuellement questionnées, nous avons construit cinq catégories :

- « rapport au référentiel » ;
- « expression de l’objet » ;
- « élaboration des tâches évaluatives » ;
- « pratiques de pondération » ;
- « pratiques de notation ».

Par ailleurs, ce travail à l’aide des catégories engage le processus de traitement des questions de recherche ancrées dans les phénomènes observés. Nous retrouvons ici aussi la dialectique inductive-déductive décrite plus haut entre un mouvement descendant, dans lequel la catégorie éclaire une compréhension du phénomène, et un autre ascendant, où le phénomène donne à voir la puissance explicative de la catégorie, mais aussi ses limites. En effet, d’un point de vue compréhensif, la catégorie ne peut à elle seule offrir une compréhension complète et exhaustive du phénomène qu’elle met en évidence conceptuellement. Il s’agit donc de croiser les catégories, de les envisager comme un système compréhensif et non comme des cellules de sens isolées. Les opérations d’analyse ne sont dès lors pas que « basement » techniques, mais relèvent d’articulations, de mises en lien conceptuelles, au travers d’un travail de questionnement des données. Pour le chercheur, l’enjeu ici est de convoquer des catégories conceptualisantes certes en rapport avec l’étude de l’objet de recherche, mais aussi, et peut-être surtout, qui vont permettre de dégager un sens conceptuel et pratique. Ce qui nous amène à les caractériser.

5.1.2 La catégorie : un condensé de significations

Le centre de l'activité de catégorisation est l'interprétation/conceptualisation. Une catégorie renvoie donc à l'articulation d'autres catégories, dans une perspective systémique, et dans une finalité de production de sens. Cela nous pousse à être en contact avec l'entier de nos données, la catégorisation n'étant pas en mesure d'être saisie qu'à l'aune des phénomènes issus du terrain. Les catégories donnent ici à voir les expériences vécues, quelles qu'elles soient, comme autant d'essais de compréhension. Elles s'enracinent dans des discours, des traces, des mots et les propulsent à un niveau synthétique de compréhension globale. Comme le dit Perrenoud (1988) : « la catégorie permet de construire une représentation théorique de certains types de pratiques, de fonctionnement, de processus, en prenant en compte ce qui leur donne sens dans l'esprit des acteurs » (p. 102).

5.1.3 La catégorie : pour l'étude de phénomènes divers

Les phénomènes que nous cherchons à observer, décrire et conceptualiser sont d'ordre évaluatifs et relatifs aux pratiques réelles des enseignants engagés. Nos catégories visent à saisir la complexité de la pratique en question de manière située et empiriquement fondée, tout en tenant compte de l'intention de communication des enseignants. De ce fait, nos catégories sont denses, parce qu'elles se rapportent à des logiques et des processus intimement ancrés dans le rapport qu'entretiennent des pratiques évaluatives avec des éléments théoriques. Elles donnent à voir ce qui se passe, ce qui est en mouvement, ce qui est en jeu dans ces pratiques, tout en étant intimement liées à nos questions de recherche (Raab, 2015).

5.1.4 La catégorie : pour expliciter la dynamique des concepts

Si nos catégories s'ancrent directement dans notre modèle théorique, elles n'en sont pas pour autant une transposition directe à l'interprétation des phénomènes observés. Ainsi, ces phénomènes ne prennent toute leur ampleur que lorsqu'une tentative de compréhension met en synergie, par exemple, la catégorie « pratiques de pondération » avec celle de « expression des objets », ou « pratiques de notation » avec « élaboration des tâches évaluatives ». Nous pourrions dire que dans cette perspective, nos catégories ont fait l'objet, dans la phase de recension des résultats, d'un travail de mise en lien, réflexif, progressif (Rondeau & Paillé, 2016).

5.1.5 La catégorie : pour interpréter toutes les données

Dans une perspective qualitative, le travail d'analyse à l'aide de catégories conceptualisantes peut se réaliser avec tout type de données. C'est bien ce que nous nous proposons de faire, au niveau de nos quatre corpus. Dans un premier temps, cette manière de procéder nous permettra de croiser l'analyse de nos données dans une perspective « horizontale » afin de dresser des portraits de chaque enseignant. Puis, dans un second temps, elle nous offrira la possibilité d'investiguer un axe d'analyse plus « vertical », en approfondissant la question des pratiques de notation et du développement des compétences évaluatives y relatives chez tous les enseignants¹²².

5.2 Catégorie et catégorisation

Dans notre méthodologie, la grille catégorielle s'inspire de notre modèle théorique de l'alignement curriculaire élargi. Il ne demeure pas moins qu'un travail de double questionnement est à l'œuvre : un questionnement des données à l'aune du modèle théorique, et dialectiquement, un questionnement du modèle au regard des données. Cette articulation est nécessaire à la compréhension des phénomènes en jeu. Là, nous retrouvons la complémentarité des mouvements inductifs et déductifs décrits plus haut. L'analyse des données ainsi conçue permet d'éprouver notre système interprétatif - le modèle théorique de l'alignement curriculaire élargi - en référence à la réalité de la pratique, mais aussi, et de manière conjointe, d'interpréter les données empiriques d'un point de vue conceptuel. Dans ce sens, nous suivons Paillé et Mucchielli (2012), lorsqu'ils affirment qu'« en acceptant les concepts comme décrivant indéniablement la “réalité” à construire, on accepte implicitement tous les principes présumés du système interprétatif de départ puisque celui-ci explique les phénomènes avec ces concepts » (p. 332). Ce positionnement interprétatif de recherche court-il dès lors le risque, comme le suggèrent les auteurs, d'être relativement clos ? Nous ne le pensons pas. En effet, et comme nous l'avons précisé dans la section précédente, les catégories donnent à voir davantage une circulation de sens au travers du processus de catégorisation, plutôt que des explications étiologiques centrées sur des épiphénomènes éclairés par des concepts isolés. Nous faisons l'hypothèse que c'est aussi et surtout grâce aux processus de triangulation des données que cette richesse de sens peut exister.

¹²² Ces éléments sont approfondis plus loin.

5.3 Un travail de déduction et d'induction, encore...

Notre démarche méthodologique est en accord avec, au départ, une observation empirique des phénomènes évaluatifs, des logiques en jeu dans les pratiques, des questions ou problèmes évoqués par les praticiens, et d'un essai de leur conceptualisation. Ce processus relève de procédés « d'induction théorisante » (Paillé & Mucchielli, 2012), où le chercheur tente d'amener les phénomènes à un niveau d'intelligibilité pour qu'ils deviennent objets de recherche. Comme nous l'avons dit, ce travail reste principalement inductif, dans la mesure où la catégorie est construite, et non reproduite d'une catégorie pré-existante. L'enjeu de ce travail consiste à élaborer un concept, et non à l'importer de manière décontextualisée. Puis, parce que tout processus d'analyse en émergence engendre des « coadaptations des observations et des théories » (Gil, 1998, p. 137), l'effort inductif va progressivement laisser la place à un travail plus déductif ancré, ici, dans un modèle théorique permettant d'interpréter les phénomènes émergents. En effet, ce cas de figure est particulièrement présent dans des recherches comme la nôtre, où les catégories, comme autant de référents, proviennent de modèles théoriques en lien avec l'objet de recherche. Là, à mesure que l'analyse monte en généralité, le chercheur doit « combiner la seule induction avec des propositions théoriques établies » (Kaufmann, 2008, p. 227).

6 L'analyse des données : une nécessaire triangulation

Nous avons déjà évoqué la triangulation comme processus analytique clé de notre méthodologie de recherche. La définition du concept de triangulation en recherche qualitative est attribuée aux travaux de Denzin (1978) et plus récemment à ceux de Silverman (2009). Ce dernier auteur souligne l'impossibilité d'obtenir des informations complètes et exhaustives sur un objet ou un phénomène étudié lorsque ces dernières émanent d'une seule source. Dans ce sens, trianguler revient à « utiliser de manière combinée différentes techniques de récolte ou d'analyse de données » (Berger, Crescentini, Galeandro & Crohas, 2010, p. 2), ou à combiner les données issues de plusieurs instruments, compensant ainsi les biais propres à chacun (Savoie-Zajc, 2009). Ce constat souligne la nécessité d'utiliser divers instruments, d'exploiter différentes données pour tenter de saisir des dimensions variées de l'objet investigué dans toute sa complexité : « two or more qualitative methods may be used sequentially or simultaneously, provided the analysis is kept separate and the methods are not muddled » (Denzin, 1978, p. 66).

Le but de la triangulation, au sens de ce chercheur, est d'arriver à identifier la valeur la plus valide des données analysées.

Dans ses travaux fondateurs de 1978, Denzin met en évidence trois types de triangulation : l'exploitation de sources de données variées pour le temps, l'espace et les personnes ; le recours à des chercheurs différents ; l'utilisation simultanée de méthodes d'analyse variées. Il différencie donc la triangulation des sources d'information, la triangulation des participants et la triangulation de méthodes¹²³. Adopter la triangulation est donc un choix méthodologique stratégique, ayant pour objectif de contrôler les biais éventuels d'interprétation dans une recherche de validité : « using only one research strategy – for example, only observation, or worse, only one kind of observation – introduces bias into the data record » (Graue & Walsh, 1998, p. 102). Dans ce sens, la triangulation conduit vers la validité (justesse et stabilité) des analyses effectuées (Baribeau & Royer, 2012), et devient intimement liée à la scientificité de la démarche méthodologique. Elle vise à conférer aux démarches qualitatives également de la rigueur, de l'ampleur et de la profondeur à la recherche (Denzin & Lincoln, 1998)

6.1 Une stratégie de recherche particulière

La triangulation est avant tout une *stratégie inductive* de recherche (Flick, 1998), dans le sens où elle privilégie et valorise le fait de partir d'un phénomène particulier observé sur le terrain pour le décrire et le comprendre. Elle se donne comme finalité d'élaborer un savoir pertinent sur le phénomène, à partir des différentes opérations de croisement des données considérées comme porteuses de sens. La triangulation peut dès lors être assimilée à une procédure de va-et-vient problématisé et dynamique entre opérations de recueil, d'analyse et de conceptualisation. Nous retrouvons ici les caractéristiques de l'analyse des données au travers des catégories conceptualisantes explicitées plus haut, et le rapport de cohérence que cette analyse entretient avec nos données qualitatives : les écrits, les entretiens, les échanges lors des actions de formation ou encore les évaluations sommatives réalisées par les enseignants ainsi que leurs discussions. Ici, le fil rouge méthodologique des processus de triangulation mis en œuvre est constitué par un aller-retour constant entre le « terrain » et l'« analyse », basé sur le principe de réinjection, prenant lui-même la forme de questionnements, d'hypothèses et

¹²³ Nous préférons ce terme à celui de « triangulation méthodologique », qui fait notamment référence aux « méthodologies mixtes » ou « *mixed-methods research* » chez les anglo-saxons, qui appartiennent stratégiquement des données qualitatives et quantitatives de façon cohérente afin d'enrichir les résultats de la recherche, ce qui n'est pas notre cas.

d'observations, de manière à réaliser de nouvelles opérations empiriques visant à construire une théorie validée à propos du phénomène étudié.

6.2 Une triangulation de méthodes

Nous nous inscrivons pleinement ici dans une triangulation de méthodes, dans la mesure où nous croisons différentes méthodes de recueil de données. En effet, nos analyses ont été réalisées en croisant quatre types de données, éclairant chacune le même objet et visant à expliciter des phénomènes récurrents. Chaque source de données, dans cette perspective analytique, apporte des informations nouvelles permettant d'augmenter les possibilités de triangulations porteuses de sens (Lincoln & Guba, 1985), et ainsi participer à la validité de la démarche.

6.3 Des perspectives de triangulation poursuivant des buts différents

Mottier Lopez et Allal (2010), fortes de diverses expériences en recherche exploitant des processus de triangulations variés, montrent que la littérature met en avant deux perspectives de triangulation : une perspective de validation, et une perspective d'élargissement de la compréhension des phénomènes étudiés.

6.3.1 Réduire les biais associés à chaque méthode

Au sujet de la première, et en convoquant de multiples travaux faisant référence, elles précisent que :

La première perspective a un but de validation, considérant qu'elle « augmente à la fois la validité et la fiabilité des données d'évaluation » (Patton, 1990, p. 245). Elle permet de réduire les biais associés à chaque méthode d'étude. Miles et Huberman (1994) parlent, dans cette même visée, de « réduction de l'incertitude », une interprétation étant validée par un faisceau d'indications concordantes. Cette perspective de validation et de vérification est qualifiée de triangulation « restreinte » par Van der Maren (1999) et de « triangulation normative » par Cohen et Manion (1994) qui insistent quant à eux sur le fait de tester la concordance entre indicateurs. (p. 240).

Cette perspective participe à renforcer la validité des résultats produits. Il aurait dans notre cas été très risqué de n'exploiter qu'une source de données, l'objet de notre recherche étant très complexe. Nous avons donc utilisé cette triangulation dans une perspective de validation des résultats, mais pas exclusivement. À notre sens, cette perspective, bien que tout à fait cohérente

avec la recherche qualitative et notre dispositif, comporte un risque : s'éloigner de la complexité de la pratique des acteurs, ici, évaluative, et ainsi perdre des informations importantes participant à mieux comprendre les phénomènes étudiés. Nous avons donc été amené à adopter dans cette recherche une perspective de triangulation « mixte »¹²⁴, s'appuyant également sur une recherche de compréhension des phénomènes. Les deux perspectives sont complémentaires. La première n'empêche pas une « recherche de compréhension » des phénomènes. Elle s'y inscrit aussi, mais dans le but de valider des résultats d'analyses en compréhension. La deuxième est utilisée pour approfondir cette compréhension.

6.3.2 Comprendre en profondeur

Il nous a semblé également pertinent d'adopter, de manière conjuguée à la perspective de validation, une triangulation à perspective compréhensive. Cette triangulation, comme le formulent les chercheuses, vise « un élargissement de la compréhension des phénomènes étudiés » (Mottier Lopez & Allal, 2010, p. 240). À nouveau, en se référant à de multiples sources, elles montrent que :

Lorsque les démarches d'une triangulation ne sont pas concordantes, Patton (1990) propose que le chercheur tente de comprendre pourquoi et sous quelles circonstances les différences se produisent. Miles et Huberman (1994) suggèrent de rechercher et de réunir des informations divergentes qui peuvent amener le chercheur à élaborer des explications plus complexes ou à adopter des nouvelles pistes de réflexion sur la signification de ses données. Van der Maren (2006) parle ici de triangulation « élargie », en tant que stratégie d'« exploration » qui a pour but d'enrichir les informations à disposition par la recherche de sources de nature différente ; il s'agit d'augmenter la « pertinence » des interprétations par « la prise en compte de la complexité du terrain et des perspectives complémentaires qui peuvent s'y exprimer » (p. 72). Quant à Cohen et Manion (1994), ils définissent une triangulation dite « interprétative » qui confronte et met en relation des visions différentes d'un phénomène multidimensionnel (p. 240).

Parce que nos données nous ont rapidement permis d'être confronté à des logiques d'acteurs décrites comme non linéaires et porteuses de tensions, et parce que notre dispositif se voulait au départ être au service d'une évolution possible des pratiques, une double triangulation s'est avérée nécessaire. En effet, il nous est apparu fondamental d'élargir chacune des perspectives de départ, en cherchant conjointement à valider nos résultats et à en étendre la portée compréhensive.

¹²⁴ A ne pas confondre avec la méthodologie mixte que nous avons évoquée précédemment.

6.4 Une première phase d'analyse à approfondir

Un des critères de validité de notre méthodologie réside, selon nous, dans le fait que notre situation de recherche induit, plus ou moins explicitement, de nous référer à un modèle théorique ayant un rapport étroit avec notre objet d'étude. Les pratiques évaluatives, dans un sens large, ont été étudiées au travers de nombreux dispositifs de recherche à l'aune de l'alignement curriculaire¹²⁵. Une littérature conséquente existe sur le sujet¹²⁶. Malgré le fait que certains espaces de recherche sur cet objet soient encore peu explorés, nous disposons donc déjà de référents valides pour penser les phénomènes évaluatifs. Comme nous avons essayé de le montrer, ces constats infléchissent l'orientation méthodologique de l'analyse de nos données, mais ne prédisent en rien ce qu'elles vont dévoiler : le travail de catégorisation reste original et spécifique à notre problématique et au contexte de la recherche, l'articulation induction/déduction répond à des questions de recherche précises, et le travail de théorisation ne s'appuie pas sur la seule transposition de concepts.

Cela dit, nous avons jugé que le travail d'analyse ne pouvait s'arrêter à une explicitation de phénomènes, même complexes et pertinents, caractérisant des fonctionnements individuels et transversaux. Nous avons par ailleurs constaté que notre démarche d'analyse à l'aide des catégories conceptualisantes, si elle était porteuse pour nos deux premières questions de recherche, s'avérait incomplète pour la troisième qui investigate les pratiques de notation des enseignants. Il nous est donc apparu nécessaire, pour traiter de ces dernières, de construire un outil d'analyse pour observer le développement de ces compétences. Pour cela, nous nous sommes inspiré de la perspective théorique de *l'assessment literacy*.

6.5 Une seconde phase d'analyse interprétative et dynamique

Les dimensions conceptuelles de *l'assessment literacy* sont développées dans la partie 4. Nous ne les reprenons pas ici. Toutefois, il nous semble important de préciser comment nous avons construit notre outil d'analyse, de manière à mettre en évidence ce qui relève pour nous d'une validité méthodologique.

Notre question de recherche B investigate les processus de développement des compétences à évaluer des enseignants, la C l'évolution des pratiques de notation. De manière à décrire ces

¹²⁵ Pour les références de ces travaux, se référer à la partie 3 traitant du cadre conceptuel.

¹²⁶ Voir la revue de littérature en partie 2.

processus de manière dynamique, il nous a semblé pertinent de montrer dans quelles mesures les catégories conceptualisantes éclairaient les différentes dimensions des compétences en évolution. Pour cela, nous avons croisé nos cinq catégories avec les six compétences à évaluer élaborées au regard des apports de la recherche, des questions des praticiens et de leur contexte, dans une démarche elle aussi inscrite dans une perspective inductive/déductive.

Pour rappel, les six compétences sont :

- A. Analyser le plan d'études du point de vue des objectifs et des objets de savoir disciplinaires qui y circulent à des fins d'évaluation sommative en mettant en relation de manière cohérente les différents niveaux de formalisation des objectifs, et en analysant les objets de savoir pour en déterminer les caractéristiques essentielles à des fins d'évaluation sommative.
- B. Elaborer des tâches évaluatives complexes en rapport avec les objectifs des plans d'études qui donnent des preuves d'apprentissages maîtrisés ou en voie de réalisation, en privilégiant des tâches qui donnent à voir des unités d'apprentissage significatives (approche qualitative).
- C. Adopter des systèmes de pondération permettant de mettre en évidence le poids donné à l'apprentissage référé aux objectifs et aux objets de savoir en privilégiant, tant que possible, une pondération critériée, notamment pour permettre un feedback aux élèves à visée de régulation.
- D. Construire des notes de manière à ce qu'elles rendent compte des acquis et lacunes des apprentissages des élèves en élaborant ses propres échelles et barèmes en référence au système de pondération, à la complexité des tâches et aux objectifs évalués.
- E. Penser ses pratiques évaluatives sommatives en terme de validité et de cohérence d'alignement curriculaire élargi en les référant aux processus d'enseignement et d'apprentissage, et en analysant leurs cohérence à l'aune d'éléments taxonomiques.
- F. Justifier ses choix en matière de pratiques évaluatives sommatives au regard de son contexte institutionnel cantonal, local (établissement) et de la culture des équipes dans lesquelles elles s'inscrivent, en les adaptant aux prescriptions institutionnelles, tout en étant attentif à ce qu'elles soutiennent les apprentissages des élèves.

Le tableau qui rend compte de ce croisement prend cette forme :

Tableau 4 : La grille d'analyse croisant les cinq catégories conceptualisantes et les six compétences

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>	<i>Catégories</i>
Compétences						
A	√	√	X	X	X	
B	√	√	√	X	X	
C	√	√	X	√	X	
D	√	√	√	√	√	
E	√	√	√	√	√	
F	√	√	√	√	√	

Le signe (√) dénote un rapport étroit entre la catégorie et les connaissances et les ressources induites dans la compétence en question. En cela, nous avons un indicateur décrivant à quel niveau une appropriation est en cours. La dimension dynamique du développement des compétences est par ailleurs visible au niveau des multiples catégories qui sont en jeu dans une compétence, mais aussi dans les liens qui peuvent être faits entre les catégories comme autant d'indicateurs de développement. Le signe (X) montre l'absence de rapport entre la catégorie et les dimensions de la compétence.

Le rapport entre développement de compétences et développement professionnel est ici potentiellement perceptible. N'étant pas directement lié à notre objet de recherche, mais entretenant des rapports étroits avec notre méthodologie ainsi qu'avec notre perspective théorique de l'*assessment literacy*, nous débattons de son statut dans notre travail lors de notre discussion finale.

7 Deux parties pour exposer les résultats

En cohérence avec notre méthodologie et nos questions de recherche, nous présenterons nos résultats en deux temps :

- Partie 7 : des portraits de six enseignants dans une perspective horizontale (Questions A et B). Nous avons ici souhaité mettre en évidence les processus d'appropriation de chacun d'entre eux. Deux enseignantes, une par discipline, ont été écartées : une en français, à cause de son faible investissement dans les séances de formation qui ne nous ont pas permis d'accéder à certaines de ses réflexions dans l'action ; et l'autre en mathématiques, par le fait qu'elle n'a pas produit de récit de pratique.

- Partie 8 : une analyse de leurs pratiques de notation, dans une perspective davantage verticale (Question C). Cette phase s'est imposée car elle rend compte d'un axe de problématisation et de questionnement fort commun à chaque enseignant.

Partie 7 - Premier chapitre de résultats : six portraits d'enseignants

1 Introduction

Ce chapitre de résultats est centré sur nos deux premières questions de recherche et donne à voir :

- Relativement à la question A, à quel.s niveau.x se situe.nt les problèmes rencontrés et exprimés par les enseignants dans leurs pratiques évaluatives sommatives et dans quelle mesure ils peuvent être dépassés grâce à leur appropriation du modèle théorique de l'alignement curriculaire élargi. Par exemple, quand les enseignants expriment un questionnement sur ce qu'est une évaluation « juste », que devient ce terme mis à l'épreuve du modèle théorique ?
- Relativement à la question B, le processus d'appropriation des connaissances et le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi, travail aux prises également avec des dilemmes, des controverses, des difficultés (Mottier Lopez & Pasquini, 2017).

Ce chapitre s'organise autour de la description de chaque parcours d'enseignant sous forme d'un portrait, en suivant les trois temps de la formation : avant, pendant et après. Ce choix nous semble pertinent à au moins trois titres :

- Il respecte notre méthodologie de recherche à visée compréhensive. En mettant en évidence le cheminement de chaque enseignant dans une perspective longitudinale et dynamique, nous serons à même de trianguler nos quatre types de données ;
- Cette méthode nous permettra de mettre en évidence les connaissances en construction, les mouvements dans les pratiques, les changements, mais aussi les résistances, les récurrences d'un certain nombre de questions, problèmes et leurs résolutions quand elles existent ;
- Enfin, ce choix nous donnera l'occasion d'éprouver notre cadre compréhensif élaboré au travers d'un temps d'analyse interprétative à l'aide de catégories conceptualisantes (Paillé & Mucchielli, 2013), et d'un temps de modélisation avec les apports de *l'assessment literacy* (e. g., Popham, 2011).

Afin de rester proche de la logique des enseignants, nous respecterons la manière avec laquelle ils ont restitué les pièces qui nous ont servi de données durant la recherche collaborative. En

effet, certains d'entre eux ont par exemple exploité les écrits pour décrire et analyser leurs pratiques évaluatives en jeu dans les épreuves discutées en séminaire, alors que d'autres y ont plutôt trouvé un lieu de réflexion plus global, ou un espace de mise à distance de questions plus identitaires. Durant la formation, des textes nous sont parvenus parfois avant la deuxième séance de formation, parfois après. Il en va de même pour les évaluations sommatives qu'ils ont été amenés à produire : certains ont travaillé individuellement, d'autres en duo, d'autres seul puis avec un collègue. Ce sont, à nos yeux, autant d'éléments à considérer comme relevant de choix porteurs de sens, comme des critères aussi de validité d'une recherche collaborative.

Ce chapitre commencera avec les portraits des enseignants de mathématiques : Fernand ouvrira la marche, puis Paul et enfin Philémon. Comme ce dernier est le seul à n'avoir pas produit d'écrit, il termine les portraits de sa discipline. Suivront les portraits des trois enseignantes de français. Nous commencerons en proposant un appariement des portraits d'Anouck et d'Alizée. Ayant collaboré de bout en bout dans la recherche, et enseignant en duo dans leur établissement, elles ont produit les mêmes épreuves. Cela ne nous empêchera pas de montrer leur cheminement respectif, malgré de nombreuses convergences. Le dernier portrait sera celui de Samia. Une conclusion, avec quelques pistes de réflexion et surtout de discussion, clôturera le chapitre.

Enfin, quelques éléments aidant à la lecture du chapitre, que nous ne répéterons pas dans chaque portrait, seront à considérer :

- Les enseignants ne connaissaient pas le modèle théorique de l'alignement curriculaire élargi avant d'entrer dans le dispositif de recherche ;
- Les nombreux acronymes auxquels ils se réfèrent sont explicités dans l'annexe Partie 1, 1 ;
- Les épreuves réalisées, reproduites partiellement dans ce chapitre, se retrouvent en intégralité dans l'annexe Partie 4.

2 Portrait 1 : Fernand, enseignant de mathématiques

La réalisation du portrait de Fernand s'appuie sur les données suivantes :

- Le premier entretien et le premier écrit, réalisés tout deux avant la formation ;

- Sa première et seconde évaluation¹²⁷ ainsi que leur discussion proposée par le formateur, son deuxième écrit, tous produits durant la formation ;
- Les échanges entre Fernand, ses pairs et le formateur durant les séminaires de formation 2 et 3.
- Le second entretien et le troisième écrit, réalisés à la fin de la formation.

Abordons maintenant les dimensions interrogées par nos questions et sous-questions de recherche.

2.1 Les dimensions des pratiques évaluatives de Fernand sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.

Avant la formation

2.1.1 Les dimensions de la pratique évaluative sur lesquelles Fernand porte son attention *avant la formation*

Fernand explique, au cours de son entretien et dans son écrit, que son problème réside essentiellement dans la pondération : « *Dans une évaluation, quel poids donner, je veux dire, sur chaque partie, sur chaque objectif, ou sur chaque... que j'évalue. Pour moi, c'est le plus difficile* ». Pourtant, il est au clair sur sa manière de concevoir ses évaluations, sur les éléments qu'il juge importants, à savoir :

- Les contenus mathématiques, qu'il appelle « *des gestes ou outils mathématiques* » ;
- Le lien étroit existant entre son enseignement et l'évaluation : « *j'ai déjà une idée du test au début du thème, en tout cas, de là où j'aimerais aller* » ;
- Le contenu et la progression des élèves : « *ça veut dire que pour chaque thème, j'ai décortiqué tout ce que l'élève, tous les gestes que l'élève doit savoir faire* ». Il évalue en test significatif essentiellement des contenus complexes, même si « *y a toujours quand même une partie technique dans le TS, mais elle est vraiment résiduelle* ».

Quatre éléments caractérisent la manière avec laquelle Fernand élabore les évaluations sommatives. Nous les reprenons et les mettons en relation avec le discours de Fernand :

¹²⁷ Fernand a réalisé seul les deux épreuves, travaillant dans un autre établissement que ses trois collègues de discipline.

- Les exercices sont en rapport avec l'enseignement dispensé. Ils ne permettent pas aux élèves de « *répondre au hasard* ». Ils « *sollicitent leur intelligence et les outils qu'ils ont acquis* ». Lorsqu'il explicite sa pratique de *design*¹²⁸ des tâches, il précise : « *Déterminer à l'aide des moyens d'enseignement un lot d'exercices types qui pourraient figurer dans le test. Ceci afin de les faire travailler, en trier les enseignements et aider les élèves à identifier les outils mathématiques nécessaires à leur résolution* ».
- Les objectifs « *font référence au PER* » et sont posés sur l'évaluation, traduits, par souci de transparence. C'est un travail important à ses yeux, car il met en évidence les éléments clés qui seront évalués. Toutefois, le PER est aussi une contrainte. Là, il critique l'organisation des contenus : « *ces découpages en degrés, en années, etc., ils vont pas toujours dans le sens de ce que j'ai appris en maths et de ce que comme on peut utiliser les ressources des élèves et le temps, c'est surtout ça* ». Malgré ces limites, dans son premier texte, il explique qu'il « *entre* » par le Plan d'études pour déterminer ce qu'il y a à évaluer lorsqu'il élabore une épreuve : « *Comme je sais qu'il peut y avoir des différences dans les progressions d'apprentissages entre les VG et les VP je commence par un passage par le PER. Ce passage me permet de modifier et de compléter ce que j'ai imaginé faire travailler aux élèves* ». Suit une réflexion sur l'importance des contenus mathématiques dans le parcours des élèves : « *Je sais par expérience que les propriétés des figures revêtent une importance particulière dans la suite du programme de VP* ». Ce travail s'accompagne de l'identification et du choix des éléments de progression des apprentissages évalués.
- La pondération est pensée au regard des apprentissages : « *Identifier pour le test quelles compétences seront génératrices de points* ». Par ailleurs il recherche à donner le même poids à chaque partie, même si le tout sera noté à l'aune d'un barème préexistant : « *Il ne restera plus qu'à vérifier l'équilibre entre les différentes parties. Pour cela je n'ai pas de critères définis. Je me base sur mon feeling/expérience et mon intuition. L'échelle sera le barème fédéral (nbre de pts atteints / nbre de pts total x 5) + 1* ».
- L'échelle est toujours présente en amont : « *y a l'échelle complète, ça veut dire, de tant de points à tant de points ben ça fait 6, de tant de points à tant de points ça fait 5,5* ». Cette pratique semble stable : « *généralement je bouge pas l'échelle. [...] Y a pas de raison qu'elle bouge* », comme utiliser des seuils fixes, ou des échelles standardisées : « *j'utilise tout bêtement ibaremes, pis comme c'est justement une accumulation de*

¹²⁸ Nous parlons d'élaboration de tâches au niveau de nos catégories conceptualisantes.

points, ben je mets, j'utilise le barème fédéral [...], car il faut en utiliser un ». Toutefois, Fernand exprime que pour lui, le barème doit être en rapport avec l'évaluation, car en tant que tel, sa signification est limitée : « pour moi, ces barèmes ils n'ont pas beaucoup de sens [...] j'entends des collègues qui me disent « ah ouais, non mais en VP, il faut que ce soit au moins 75% ». Et pour moi, ça a pas de sens, parce que c'est lié au test ». Il ne s'autorise pas à bouger les seuils pour les TS, pour lesquels « c'est toujours le barème fédéral ». La raison invoquée est le nombre de points : « Sur les TS, on est sur quelque chose comme quarante points, cinquante points, donc, je veux dire, l'échelle elle se répartit, les pourcentages ils sont plus harmonieux, je dirais, de façon naturelle, mathématique ». Cela dit, le nombre de notes imposées par le Département a une influence : « ce nombre de notes contraint ma liberté d'évaluateur ».

Nous observons qu'avant d'entrer en formation, Fernand semble au clair sur des éléments de sa pratique qu'il juge comme stables et d'autres qui le questionnent davantage. Nous voyons que ces derniers se situent principalement au niveau de la pondération, mais nous faisons ici l'hypothèse qu'ils concernent aussi la notation, puisque des tensions existent dans son rapport aux barèmes.

Pendant la formation

2.1.2 Les points de cohérence et de ruptures d'alignement observés par Fernand dans ses pratiques d'évaluation sommative lorsqu'il se réfère au modèle de l'alignement curriculaire élargi *pendant la formation*

Rappelons ici brièvement les trois temps de la réponse à cette sous-question :

- Présentation des données brutes en nous centrant sur les parties qui ont suscité un débat ou qui mettent en évidence des zones de tensions, de discussions, et en respectant l'ordre chronologique dans lequel elles ont été produites (ici : épreuve 1, épreuve 2, écrit 2).
- Présentation d'un tableau de synthèse de ces éléments que nous interpréterons en référence aux catégories conceptualisantes que nous avons définies.
- Commentaires de nos résultats.

2.1.2.1 Première évaluation créée pendant la formation

Cette évaluation porte sur les isométries et les propriétés des polygones. Elle se destine à des élèves de 9e VP (Voie Prégymnasiale) et n'a pas encore été mise en œuvre au moment où elle

est présentée et discutée. L'objectif du PER est mentionné au travers de ses composantes. L'échelle est posée en amont avec des seuils pour chaque note.

	Nom :
	Prénom :
	Date :
	Classe :

Mathématique

Evaluation significative

Géométrie

MSN 31 - Poser et résoudre des problèmes pour modéliser le plan et l'espace ...

... en définissant des figures planes par certaines de leurs propriétés géométriques ;

... en utilisant des propriétés des figures et leur décomposition en figures élémentaires pour les construire et les reproduire ;

... en utilisant les instruments appropriés ;

... en mobilisant des transformations géométriques

Echelle :

40 à 38.5 6	28 à 25 4	14.5 à 12 2
38 à 35 5.5	24.5 à 22 3.5	11.5 à 8.5 1.5
34.5 à 32 5	21.5 à 18.5 3	8 à 0 1
31.5 à 28.5 4.5	18 à 15 2.5	

Remarques :

.....

.....

Nota bene.

Pour les 3 premiers exercices il y a un point par exercice pour le respect de la consigne.

Pour chaque dessin il y a un point pour la précision et le soin apporté à sa réalisation

Figure 7 : Fernand, épreuve 1, géométrie, page de garde

L'épreuve est constituée de quatre tâches. Les points figurent à côté de chacune d'entre elles. La manière avec laquelle ces derniers sont attribués est précisée dans la table de spécification qui suit (Figures 6 et 7). Les éléments de discussion relevés par le formateur¹²⁹ sur cette page

¹²⁹ Rappelons que nous sommes le formateur. Pour une discussion critique sur son rôle dans cette recherche, le lecteur se dirigera vers la conclusion finale de ce manuscrit.

sont les suivants¹³⁰ :

- Un questionnement sur la pertinence d'évaluer la deuxième composante de l'objectif (... en utilisant des propriétés des figures ...) à hauteur de 3 points ;
- Une interrogation sur la convocation de composantes du PER plutôt que d'éléments de progression des apprentissages ;
- Le fait que, dans l'échelle, des critères soient appariés à des points, et que ce sont ces derniers qui servent à déterminer les seuils de chaque note.

1. a) Construis 2 triangles équilatéraux ABC et DBC dont $AB = 5 \text{ cm}$: /14 pts
 b) Comment s'appelle le quadrilatère ABCD ?
 c) Quelles sont ses propriétés ?

 d) Construis $A'B'C'D'$ image de ABCD par symétrie centrale de centre C.
 e) Combien de côté compte le polygone ABDA'B'D' ?
 f) Sais-tu comment appelle-t-on un polygone avec le nombre de côtés que tu as trouvé en e) ?
 g) Comment s'appelle le quadrilatère ADA'D' ?
 h) Quelles sont ses propriétés ?



Figure 8 : Fernand, épreuve 1, géométrie, tâche 1

Ici, le formateur questionne la distribution des 14 points.

¹³⁰ Le bien-fondé de ces éléments est présenté dans la partie 6 relative à la méthodologie.

2. . a) A' est l'image de A par une symétrie d'axe s. Retrouve l'axe s. /14 pts
 b) Trace le triangle A'B'C' dont l'angle $\alpha = 110^\circ$ A'B' = 5 cm et A'C' = 8 cm
 c) Comment s'appelle ce triangle ?.....
 d) Trace A''B''C'' image de A'B'C' par R (A ; -70°)
 e) Trouve A'''B'''C''' image de A'B'C' par une translation de vecteur AA'

A
x

A'
x

Figure 9 : Fernand, épreuve 1, géométrie, tâche 2

3. a) Trouve A' qui est l'image de A par une translation de vecteur OA. /10 pts
 b) Trouve le segment A''A''' image du segment AA' par R (O ; $+40^\circ$)
 c) Construis le trapèze AA'BC isocèle dont AB = 4 cm

A
x

O
x

Figure 10 : Fernand, épreuve 1, géométrie, tâche 3

4. Pourquoi dit-on des translations, des symétries axiales et des rotations que ce sont des isométries ?.....

.....

/ 2 pts

Total/40 pts

Figure 11 : Fernand, épreuve 1, géométrie, tâche 4

Sur l'ensemble de l'épreuve, le formateur interroge Fernand sur les caractéristiques des objets de savoir évalués, en lui demandant s'ils sont selon lui tous présents.

En complément, deux « tables de spécification » accompagnent l'évaluation de Fernand. La première met en rapport les tâches ou parties de tâches avec les objectifs évalués, ce qui permet notamment à Fernand de mettre en évidence le poids donné à chaque objectif.

Evaluation TS propriété des quadrilatères et isométries					
	Ex 1	Ex 2	Ex 3	Ex 4	Total
1...en définissant des figures planes par certaines de leurs propriétés géométriques	b/c/d/e/f/g/h	c			8
2.....en utilisant des propriétés des figures et leur décomposition en figures élémentaires pour les construire...	a	b	c		3
4...en utilisant les instruments appropriés	a/d	a/b/d/e	a/b/c		9
7...en mobilisant des transformations géométriques	d	a/d/e	a/b	2	8
A.... Précision, application et propreté	a d	a/b/d/e	a/b/c		9
B... Respect de la consigne et des lettrages	1	1	1		3
Total	14	14	10	2	40

Figure 12 : Fernand, épreuve 1, géométrie, table de spécification 1

La seconde tente de faire le rapport entre les points accordés aux composantes de l'objectif MSN31, jugées plus parlantes par Fernand, et la construction de la note. Nous observons ici que, dans un premier temps, il a gradué les nombres de points référés à chaque dimension de l'apprentissage en tenant compte de leur rapport avec les tâches puis, dans un second temps, il a exploité ce travail pour déterminer le seuil de chaque note en additionnant les points.

	6	5.5	5	4.5	4	3.5	3	2.5	2	1.5	1
1...en définissant des figures planes par certaines de leurs propriétés géométriques	7/8	6/8	6/8	5/8	5/8	4/8	4/8	3/8	2/8	1/8	0/8
2.....en utilisant des propriétés des figures et leur décomposition en figures élémentaires pour les construire...	3/3	3/3	3/3	2/3	2/3	1/3	1/3	1/3	1/3	0/3	0/3
4...en utilisant les instruments appropriés	9/9	8/9	8/9	7/9	6/9	5/9	4/9	3/9	2/9	2/9	1/9
7...en mobilisant des transformations géométriques	7/8	6/8	5/8	4/8	3/8	3/8	2/8	2/8	1/8	1/8	0/8
A.... Précision, application et propreté	8/9	7/9	5/8	4/8	3/8	3/8	2/8	2/8	1/8	1/8	0/8
B... Respect de la consigne et des lettrages	3/3	3/3	2/3	2/3	1/3	1/3	1/3	1/3	0/3	0/3	0/3
Total pour comparaison ave	37	33	29	24	20	17	14	12	7	4	1

Figure 13 : Fernand, épreuve 1, géométrie, table de spécification 2

2.1.2.2 Seconde évaluation créée pendant la formation

Cette évaluation porte sur les aires et périmètres. Elle se destine à des élèves de 9e VP (Voie Prégymnasiale) et n'a pas encore été passée par les élèves au moment où elle est présentée en séminaire 3. Les objectifs sont mentionnés dans un langage traduit. L'épreuve est constituée de onze questions. Les points figurent à côté de chacune d'entre elles. L'échelle est fixée en amont et figure sur la première page.

Sur cette page, les éléments de discussion mis en évidence par le formateur sont les suivants :

- Le rapport entre les objectifs annoncés et les éléments de progression des apprentissages du PER semble peu clair ;
- Le fait que l'objectif 2 est pondéré à hauteur de 20 points, alors que les 1 et 3 ne valent respectivement que 5 points demande justification ;
- Le rapport entre les questions 1, 4a et 6a avec les objectifs évalués mérite d'être explicité ;
- L'apparition de demi-points dans l'échelle, alors qu'ils ne sont pas mentionnés dans les exercices, interroge ;
- Le fait que la réalisation de croquis et de dessins est dotée de points et que cela ne se rapporte à aucun objectif mentionné. Idem pour les phrases ;
- L'effet d'aplatissement de l'échelle, qui mériterait justification.

Mathématiques

Evaluation significative

Nom :

Prénom :

Date :

Classe :

Objectifs :

1. Je sais mesurer et calculer le périmètre d'une figure géométrique
2. Je sais mesurer et calculer l'aire des polygones
3. Je sais transformer les unités de mesure de longueur et de surface
4. Je sais trouver les dimensions d'un rectangle en connaissant les proportions de ses côtés

Echelle :

42 à 40.5	6	30,5 à 28	4	12,5 à 8,5	2
40 à 37.5	5.5	27,5 à 23	3.5	8 à 3	1.5
37 à 34.5	5	22,5 à 18	3	2,5 à 0	1
34 à 31	4.5	17.5 à 13	2.5		

Remarques :

Figure 14 : Fernand, épreuve 2, géométrie, page de garde

Pour ne pas surcharger le texte, nous ne reproduisons ici que les questions 3, 8 et 9 ayant fait l'objet d'une mise en discussion par le formateur¹³¹, discussion menée en séminaire de formation sur la base du document écrit.

<p>Question 3:</p> <p>Donne les dimensions d'un rectangle dont le périmètre et l'aire mesurent respectivement 22 cm et 30 cm².</p> <p>Note ta démarche :</p>	<p>..... / 4</p>
--	------------------

Figure 15 : Fernand, épreuve 2, géométrie, tâche 3

¹³¹ L'épreuve complète peut être consultée à l'annexe Partie 5, 9.1.2.

L'élément de discussion porte ici sur le rapport entre ce problème et les contenus évalués.

Question 8 :

Un jardin rectangulaire de $2,25 \text{ dam}^2$ est divisé en deux parties elles-mêmes rectangulaires.

L'aire de la plus grande partie est le double de la petite.

a) fais un croquis de la situation

..... /5

Figure 16 : Fernand, épreuve 2, géométrie, tâche 8

Ici, c'est l'attribution des 5 points qui est interrogée.

<p>Question 9 :</p> <p>Dans la cour de l'école, on peint un échiquier dont le côté de chaque case mesure 4 dm.</p> <p>Pour rappel un échiquier est le plateau d'un jeu d'échec. Il est constitué de 64 cases blanches et noires en alternance.</p> <p>Quelle est l'aire en m² de la surface peinte en noir ?</p> <p>Réponses :</p> <p>Quelle est le périmètre en m de cet échiquier ?</p> <p>Pour rappel un échiquier est le plateau d'un jeu d'échec. Il est constitué de 64 cases blanches et noires en alternance.</p>	<p>..... /4</p>
--	-----------------

Figure 17 : Fernand, épreuve 2, géométrie, tâche 9

Là, le formateur questionne les questions « à tiroir » en prenant comme exemple celle-ci.

Puis, un document de type « table de spécification » montre dans quels exercices et avec quelle pondération les progressions d'apprentissage du PER sont évaluées. On peut également y observer une tentative de calcul de total de points par progression d'apprentissage du PER, et

par exercice, de manière à établir des seuils pour les différentes notes.

Grille d'évaluation TS aire, périmètres & changements d'unités de longueur et de surface. MSN												
Progressions d'apprentissage	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Ex. 6	Ex. 7	Ex. 8	Ex. 9	Ex. 10	Ex. 11	Total
Comparaison, classement et mesure de grandeurs (longueur, aire, volume, angle, masse) par manipulation de lignes, angles, surfaces, ou solides, en utilisant des unités conventionnelles et non conventionnelles (1)		1					1	1	1	1	1	6
Estimation de grandeurs, choix d'une unité adéquate, prise de mesure à l'aide d'un instrument adapté et expression d'une grandeur dans diverses unités : (4, 5, 7)												0
Mesure des dimensions adéquates et calcul : (9)			1									1
du périmètre d'un polygone				1							1	2
de l'aire d'un carré, d'un rectangle, d'un triangle, d'un parallélogramme, d'un losange (par décomposition et à l'aide d'une formule)	1	1	1	1	1	1	1	C	2	1	1	12
de l'aire d'un polygone par décomposition en figures simples											1	1
tri et organisation des informations (liste, tableau, schéma, croquis...)						1	1	a			1	4
mise en œuvre d'une démarche de résolution												0
ajustement d'essais successifs												0
pose de conjectures, puis validation ou réfutation												0
déduction d'une ou plusieurs informations nouvelles à partir de celles qui sont connues					2	1		b		0		4
réduction temporaire de la complexité d'un problème												0
utilisation des propriétés des figures et des grandeurs pour établir des preuves	1	1	1	1						0		4
vérification, puis communication d'une démarche et d'un résultat en utilisant un vocabulaire, une syntaxe ainsi que des symboles adéquats			1	1	1	1		1	1	1	1	8
												0
Précision du dessin				1								1
												0
Total	2	3	4	5	4	4	3	5	4	3	6	43

Figure 18 : Fernand, épreuve 2, géométrie, table de spécification

2.1.2.3 Deuxième écrit réalisé pendant la formation

Fernand prend l'initiative de faire déjà un premier bilan de la formation. Pourtant, il écrit ce texte trois jours avant la dernière demi-journée de formation, c'est pourquoi nous respectons cet ordre-là dans la présentation des données. Les thèmes qu'il met en évidence dans ce deuxième écrit sont les suivants : un rapport au PER différent, une centration sur les contenus mathématiques, une identification des objets dans le PER, une tendance à évaluer moins de contenus.

Son récit, long d'une page et demie et présenté en annexe Partie 5, 5.2, met principalement en évidence le fait que l'alignement curriculaire élargi est désormais pour lui un outil théorique lui permettant de penser une cohérence dans ses évaluations sommatives et dans son enseignement, touchant plusieurs composantes de la pratique : « *Maintenant je peux affirmer que le PER influence non seulement le processus d'enseignement d'un thème mais il fait partie intégrante du processus d'élaboration de l'évaluation* ». Élément intéressant, le modèle lui a permis de mettre des mots sur certaines de ses habitudes relatives aux objets mathématiques : « *Cela a eu pour effet de faire le lien entre ma façon d'évaluer par « paiement au geste », liées aux prises*

de consciences que je considère comme essentielles pour la maîtrise minimale des différents objectifs fondamentaux, et les éléments de progression des apprentissages du PER ». Au cœur de ce raisonnement, le terme de progression prend alors pour Fernand tout son sens : « Travailler avec les EPA¹³² m'aide à structurer mon travail. Cela m'aide à mieux situer les élèves par rapport à là où j'aimerais les amener en lien avec les progressions que j'aimerais les voir s'approprier. Les EPA sont devenus les objets évalués. Les gestes mentaux et les prises de consciences que j'essayais d'évaluer auparavant sont devenus les outils qui vont permettre aux élèves d'acquérir les EPA sur lesquels portera l'évaluation ». Cette réflexion, qui touche le rapport de cohérence qui devrait exister entre les objectifs et les contenus évalués, influence également les pratiques de pondération de Fernand, qui peut désormais envisager d'autres manières d'attribuer des points dans ses épreuves. Par exemple, relativement à la seconde, il écrit : « dans mon évaluation 2, l'algorithme d'un calcul mal maîtrisé ou non maîtrisé n'empêchait pas l'obtention maximum des points ». On observe donc que la référence ici est l'apprentissage.

Toutefois, un objet de tensions, de débat, persiste autour de la notation : « C'est un fait, ma pratique tend de plus en plus vers un alignement curriculaire que j'essaie de rendre le plus cohérent possible sans toutefois que cet alignement soit absolu. [...] En effet, je construis l'échelle encore de manière « bâtarde ». Je fixe le 4 selon le seuil de points minimums que j'estime pour l'obtention de cette note et ce à l'aide des éléments de progression. Les autres notes naissent de ce point de référence de la façon la plus linéaire possible ».

2.1.3 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Fernand à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi pendant la formation

Sont prises en compte ici trois des quatre sources de données particulièrement significatives durant la formation :

- Les éléments de discussion proposés par le formateur (annexe Partie 5, 9). Mentionnés plus haut, ils se réfèrent tous au modèle théorique de l'alignement curriculaire élargi. Ils sont reproduits dans le tableau ;
- Les éléments débattus en séminaires de formation 2 et 3, retranscrits depuis les enregistrements ;

¹³² C'est le terme qu'utilise Fernand pour parler des éléments de progression des apprentissages du PER.

- Le deuxième écrit documenté précédemment.

Le tableau ci-dessous rend compte de la triangulation de ces données référée à nos catégories conceptualisantes.

Tableau 5 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Fernand durant la formation

<i>Données</i> Catégories	<i>Evaluation 1 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Fernand débattus en séminaire 2 de formation</i>	<i>Evaluation 2 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Fernand débattus en séminaire 3 de formation</i>	<i>Ecrit 2</i>
Rapport au référentiel	Questionnement sur l'absence de progressions des apprentissages comme « éléments de résolution de problèmes », « figures géométriques planes », ou « transformations géométriques »				Changement de statut du PER
Expression de l'objet	Question réflexive : Toutes les caractéristiques des contenus évalués sont-elles présentes ?	« s'il manque de la cohérence dans les contenus au départ, ça augmente la difficulté de mettre la note »		« je me suis posé la question de sortir la conversion d'unités pour en faire un TA, ou de dire elle est testée avant et je n'en tiens plus compte ici »	Nouvelle compréhension des objets mathématiques
					Attention accrue sur la progression des apprentissages des élèves en fonction des contenus
Elaboration des tâches évaluatives		C'est en mettant les exercices en rapport avec les composantes qu'il justifie les éléments de contenu qui peuvent être discriminantes dans la	Justification du rapport entre la question 1 et les objectifs		Centration sur leur nombre et leur niveau de complexité

		construction de la note			
Pratiques de pondération	Pertinence d'évaluer une composante de l'objectif à hauteur de 3 points sur 40 ?	Pondération du nombre de points par exercice en référence aux progressions des apprentissages. Justification au regard de l'alignement curriculaire en se centrant sur les liens entre les composantes, les exercices et les points	Pas prêt à exploiter cette démarche qualitative avec les élèves, car ne sait pas « comment l'amener ». Ce travail est utile pour voir « l'équilibre du test »	« je ne suis pas prêt à passer sous cette forme de grille critériée »	Manière d'attribuer les points en référence aux contenus
	Questionnement de la forte pondération du respect des consignes et de la précision, non référée à des contenus, à hauteur de 12 et 9 points sur 40	Il se demande « comment gérer les écarts de points entre composantes ? »	Complexité, parfois, de définir « la valeur juste » de chaque question relativement aux objectifs	La grille permet d'avoir la valeur juste d'un exercice en référence aux progressions des apprentissages	
	Interrogation de l'attribution des points en référence aux quatre tâches				
Pratiques de notation	Questionnement d'une recherche d'éléments qualitatifs qui se résout par un total de points pour construire la note	Questionnement : « comment ne pas écraser » tout ce travail qualitatif ?	Trop grande complexité à construire une note centrée sur des unités d'apprentissage	Comment différencier un 4 d'un 4,5 avec des critères ?	Difficultés à construire une note centrée sur des unités d'apprentissage : « je construis l'échelle encore de manière « bâtarde » » Echelles encore linéaires en-dessus et en-dessous du 4
		Il se dit « coincé pour faire la note »	Il dit ne pas « être prêt » à faire le pas, pour des raisons de manque d'énergie et de temps		
		Une note référée aux composantes : il peut « dire que celui qui a 6, au niveau de la composante C1, il faut ça, etc. ». Une recherche de qualité des éléments évalués.			
		Enjeux : jusqu'où il est possible de penser cette qualité dans l'évaluation et que ce soit tenable d'un			

		point de vue pragmatique ? être au fait sur les éléments incontournables nécessaires à définir les différents seuils de réussite			
--	--	--	--	--	--

Cette analyse au travers des catégories conceptualisantes met en évidence une progression dans les raisonnements évaluatifs de Fernand, car il pense désormais les pratiques et les phénomènes en mettant en lien différentes dimensions relatives à l'alignement curriculaire élargi. En effet, dans les deux épreuves et dans son écrit intermédiaire, Fernand montre qu'il accorde maintenant une attention particulière à référer les tâches aux objectifs et à concevoir l'attribution des points en rapport (voir les deux tables de spécification). La question de savoir combien de points accorder à telle ou telle composante de l'objectif du PER ou à des éléments de progression des apprentissages du PER est permanente. L'évaluation peut ainsi être lue par les objectifs tirés du plan d'études. La pondération reste parfois peu transparente pour les élèves, mais Fernand semble être en mesure le plus souvent de la justifier : « *D'ailleurs, dans mon évaluation 2, l'algorithme d'un calcul mal-maîtrisé ou non-maîtrisé n'empêchait pas l'obtention maximum des points. Je pense que cela aurait été très difficile il y a encore quelques semaines. Grâce à cette formation cela est non seulement devenu possible, mais je suis étonné de la sérénité avec laquelle j'arrive à le faire...* ». Cette pondération, sur la même épreuve, fait d'ailleurs débat lors du séminaire 3. Il va même jusqu'à penser la note en référence à ce précédent raisonnement, puisqu'il envisage des seuils en fonction des points attribués aux différentes dimensions de l'apprentissage dans les tâches. Un renversement semble s'être opéré. Avant la formation, Fernand était préoccupé déjà par cette question de pondération, mais il la considérait davantage en référence aux tâches, alors qu'elle devient progressivement significative au regard des objectifs, puisque les points ne sont plus alloués aux exercices mais aux apprentissages visés dans chaque exercice.

Egalement très préoccupé par la question des objets évalués au début de la formation, Fernand les référence maintenant au PER. Ils ne sont plus seulement présents dans ses réflexions au moment de créer l'évaluation, mais déjà au début de ses séquences d'enseignement. Il est par ailleurs intéressant de constater que cette nouvelle appréhension du contenu du PER lui permet de le considérer comme un outil de communication avec les élèves. Là, ce qu'il appelle les « *gestes mathématiques* » prennent forme au travers des éléments de progression des

apprentissages du PER.

La discussion de ses deux évaluations pendant les séminaires a surtout mis en évidence une tension entre deux manières de construire la note : la première se réfère à des unités d'apprentissage ; la seconde, bien que pensée autour d'une élaboration du seuil de suffisance également référée aux contenus, convoque *in fine* un barème linéaire. Nous retrouvons cette tension également dans son écrit, lorsque Fernand dit encore avoir recours à des échelles faites de manière « *bâtarde* », c'est-à-dire des échelles qui « *écrasent* » toute la dimension qualitative des apprentissages évalués exprimée au travers des critères en référence à l'alignement curriculaire élargi. S'il a acquis le fait que le seuil de suffisance peut être conçu en référence aux apprentissages, il demeure que construire une échelle entière dans cette perspective reste problématique pour lui. À ses yeux, le risque de déstabiliser les collègues et les élèves est grand, et il n'est pas sûr que le système dans lequel il évolue soit prêt au changement, les contraintes étant très fortes.

Un autre élément nous semble également en construction : l'éventualité d'évaluer séparément des contenus de basse complexité et ceux relevant davantage de niveaux de complexité élevée. Questionné sur ce point par Philémon lors du séminaire 3, Fernand dira : « *je me suis posé la question de sortir la conversion d'unités pour en faire un TA, ou de dire elle est testée avant et je n'en tiens plus compte ici* ». Mais le pas n'est pas encore réalisé.

Nous sommes dès lors en mesure de formuler quelques hypothèses interprétatives traduisant des mouvements dans les pratiques de Fernand, et qui nous donnent à voir dans quelle mesure il parvient à travailler la cohérence de ses pratiques évaluatives du point de vue du modèle théorique :

- Le PER devient une référence pour guider l'enseignement et circonscrire les objets mathématiques. C'est également un outil de communication des apprentissages aux élèves ;
- L'expression de l'objet de savoir est en lien avec le PER, il exprime les incontournables dans l'évaluation. Une nouvelle compréhension de cet objet est envisagée, conçue comme influençant l'évaluation, notamment au niveau de la pondération et jusqu'à la construction de la note. C'est notamment à partir de ce raisonnement que Fernand détermine le seuil de suffisance en fonction des objets mathématiques ;

- L'élaboration des tâches évaluatives devient une entrée privilégiée dans le raisonnement évaluatif, notamment au niveau du rapport étroit qu'elles doivent entretenir avec les objectifs ;
- Les pratiques de pondération appartiennent critères et points ;
- Les pratiques de notation sont pensées par rapport à des unités d'apprentissage.

Toutefois, il se dégage également des tensions, ou difficultés à dépasser des ruptures d'alignement curriculaire élargi. Les deux premières sont relevées par le formateur dans les deux épreuves que Fernand a réalisées, et restent non résolues :

- Le niveau des objectifs du PER retenus est parfois peu explicite, leur traduction à justifier ;
- La pondération des tâches reste quelques fois opaque ;
- Le recours à des barèmes linéaires est encore d'actualité, notamment pour distribuer les seuils en dessus et en dessous du seuil de suffisance (note 4) ;
- La complexité des objets peut encore être multiple au sein d'une même épreuve.

La dernière est mise en évidence par Fernand dans son écrit :

- La construction de la note sur l'ensemble de l'échelle de notes (1 à 6) dans une perspective de mise en évidence des apprentissages réalisés reste problématique, car elle est en tension avec les systèmes de pondération construits et des dimensions contextuelles (directives départementales demandant des échelles linéaires, habitudes dans les pratiques de l'établissement).

Après la formation

2.1.4 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Fernand à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi à la fin de la formation

Fernand a produit son troisième écrit avant le dernier entretien de recherche. Il l'a rédigé sous forme de bilan final. Le tableau ci-dessous rend compte des éléments significatifs de ces deux sources de données relativement aux points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre. Ils sont directement mis en rapport avec les catégories conceptualisantes, de manière à rester synthétiques et centrés sur nos questions de recherche. Certes, des redondances apparaissent, mais elles doivent être comprises et interprétées comme

des signes d'un processus complexe d'appropriation de divers éléments par Fernand. Nous commentons ces données plus bas en les mettant en dialogue avec les autres données.

Tableau 6 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Fernand à la fin de la formation

Données	Ecrit 3	Entretien 2
Catégories		
Rapport au référentiel	« lorsque je propose aux élèves d'étudier un thème, je leur propose bien souvent d'explorer au-delà du prescrit »	Découpage des contenus plus analytique
		Transparence accrue entre objectifs travaillés et évalués
		« l'attribution des points dans le travail, dans ce qui est évalué, dans les exercices, la façon d'attribuer. Avant c'était par gestes, maintenant, c'est vraiment lié aux progressions d'apprentissage » ; « comme il y a les progressions d'apprentissage, ça veut dire que c'est ça qui va pouvoir donner des points dans le travail que j'ai proposé en évaluation »
		« je me rattache aux progressions d'apprentissage et qu'avant c'était au geste, et que derrière un geste, il pouvait y avoir deux ou trois progressions d'apprentissage »
Expression de l'objet	Meilleure considération des objets d'enseignement « la volonté d'intégrer cette composante qualitative dans le processus a modifié la qualité de mon enseignement et m'a permis de lâcher certaines habitudes comme celle d'évaluer systématiquement les algorithmes de calcul dans tous les tests » « ça me permet de faire des choix en conscience » ; « c'est vraiment de dire, en fait, on est en train d'évaluer ça. Je veux dire, on est dans ce thème, on aligne avec... Et qu'à un moment, on peut pas mettre tous les thèmes... » Le seuil de suffisance est calculé en rapport avec le contenu	Une recherche de cohérence programme-enseignement-évaluation
		L'objet évalué est questionné par la manière de mettre des points
Elaboration des tâches évaluatives	Moins d'exercices, mais plus ciblés	Les exercices sont davantage pondérés en référence aux progressions d'apprentissage du PER
	« je peux mieux évaluer l'équilibre du test, des différentes parties »	« Et je vais pouvoir équilibrer, je veux dire, trouver un équilibre pour qu'il n'y ait pas trop de l'un, trop de l'autre [des progressions des apprentissages] etc. dans mon évaluation. Et ça va pouvoir me donner les exercices que je vais proposer en évaluation »
	« un exercice, maintenant, il a souvent beaucoup plus de points qu'avant »	

Pratiques de pondération	La question des demi-points tracasse encore, « <i>c'est un signe que ce que j'évalue n'est pas encore clair</i> »	La question des demi-points interroge les objets de l'évaluation
	Un exercice complexe aura plus de points	Evaluation critériée est toujours en tension
	« <i>j'ai essayé de mettre quelques critères, et puis j'ai... je les ai gardés pour moi... Comme on regarde un tableau de maître qu'on a dans son coffre (rires)</i> »	
	Difficulté de justifier la différence entre un 4 et un 4,5 avec des critères	
Pratiques de notation	Compromis dans une construction de note référée à l'apprentissage	Un « changement de référentiel » est signalé
		Construire le seuil du 4 en référence à des unités d'apprentissage n'est pas un problème
		« <i>ah lui il a 4, mais j'ai l'impression qu'il n'a rien compris</i> » [...] <i>Mais ça, j'ai l'impression que ça a disparu</i> »
		La note est toujours « écrasée » : « <i>il manque le pas [...]</i> ». « <i>Ben après, la contrainte on l'a vue, c'est des échelons en dessus et au-dessous...</i> »
		Décrire chaque note est subjectif : « <i>le barème fédéral est quand même rassurant</i> »
		« <i>je crois que c'est ça [le seuil de suffisance] qui m'importe, pour moi. Finalement, le 5, 5.5, le 6, ils m'intéressent pas</i> »

Ce tableau nous permet de formuler des hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi :

- L'utilisation du PER est critique et participe à préciser les objets de savoir ;
- Les objets de savoir conditionnent le rapport au PER et la pondération en points. Ils sont centraux dans le raisonnement sur les processus d'enseignement et d'apprentissage et les démarches d'évaluation ;
- Les évaluations portent uniquement sur des tâches complexes (problèmes), qui se réfèrent à des éléments de progression des apprentissages dont l'équilibre est recherché ;
- Les pratiques de pondération sont référées aux apprentissages ;
- Les pratiques de notation sont référées aux apprentissages, basées sur des points liés à des critères. Une importance particulière est accordée au seuil du 4 qui est aligné curriculairement, le plus possible référé à des unités d'apprentissage.

S'en dégagent également des tensions, ou difficultés à dépasser des ruptures d'alignement curriculaire élargi :

- L'utilisation de points prédomine, les critères étant trop problématiques pour noter et risqués dans les contrats avec les élèves ;
- La construction des notes « écrase » toujours le travail qualitatif réalisé en amont.

Il est intéressant de noter qu'une grande partie des réflexions menées par Fernand à la fin de la formation portent toujours sur ce qui l'a motivé à participer à la recherche : la pondération. Certes, son questionnement a évolué et s'est outillé théoriquement : *« les apports théoriques, mais on n'en a pas eu des tonnes. Je veux dire, c'était vraiment des espèces de... une colonne vertébrale »* (entretien 2, FP 18) ; *« La notion d'alignement était en partie présente chez moi bien que de façon inconsciente. La formation m'a fait conscientiser ce concept qui fait partie intégrante de ma pratique. L'AC est maintenant toujours présent dans toutes les étapes de mon enseignement. En début de la formation, je voyais bien ses mécanismes mais je pensais qu'il aurait une fonction bridante, limitante à mon enseignement »* (écrit 3). Fernand interroge plus précisément sa manière de considérer l'attribution des points dans les épreuves en référence aux objets de savoir traités. C'est d'ailleurs au travers de cette problématique qu'il pense toutes les dimensions de sa pratique évaluative. En effet, questionner la pondération lui permet conjointement d'analyser la complexité des tâches ainsi que leur nature et leur nombre, d'appréhender des modalités de notation critériées et critérielles ainsi que leurs limites, mais également de reconsidérer les objets mathématiques en jeu dans son enseignement et ses épreuves sommatives, ainsi que leur rapport au PER (épreuves 1 et 2). À ce titre, il est intéressant de constater que beaucoup de données présentées dans notre tableau font référence, plus ou moins directement, à la pondération.

Sa prise de conscience qu'une évaluation sommative peut être potentiellement critériée lui a permis de construire des outils allant dans ce sens. Il ne les a pas mis en œuvre, pour des raisons essentiellement contextuelles selon lui, mais le fait de penser l'évaluation des apprentissages d'un point de vue plus qualitatif lui a permis de remettre en question bon nombre de ses croyances et habitudes, et surtout de s'outiller pour produire des feedbacks aux élèves : *« je pense que la volonté d'intégrer cette composante qualitative dans le processus a modifié la qualité de mon enseignement et m'a permis de lâcher certaines habitudes comme celle d'évaluer systématiquement les algorithmes de calcul dans tous les tests »* (écrit 3).

Toutefois, l'écueil de la construction de l'échelle en référence aux apprentissages demeure. Le fait de convoquer un barème externe aux processus d'enseignement et d'apprentissage et

d'évaluation est toujours décrit par Fernand comme une pratique discutable même si le compromis choisi est acceptable pour lui, puisqu'il se base quand même sur un raisonnement pédagogique visant l'élaboration d'un seuil de suffisance en référence aux objectifs. Des tensions persistent donc toujours : « *J'ai conscience que j'écrase la composante qualitative par une note unique et une échelle bâtarde composée d'un seuil de suffisance qualitatif et d'échelons répartis de manière linéaire vers le haut et vers le bas* » (écrit 3). En effet, le raisonnement opéré par Fernand au niveau de la pondération bute sur des écueils importants lorsqu'il s'agit de considérer différemment une construction de note qui se voudrait, elle aussi, qualitative, c'est-à-dire référée à l'apprentissage : si poser un seuil de suffisance pour le 4 en référence aux contenus lui semble réaliste, qualifier onze notes au travers d'autant d'échelons lui paraît trop subjectif, voire dangereux, au regard des pratiques courantes. Fernand ne peut donc aller jusqu'au terme de sa réflexion. Mais ce n'est à ses yeux pas si grave, l'approche lui permettant d'améliorer grandement la cohérence non seulement de ses évaluations sommatives, mais aussi de son enseignement et de ses raisonnements pédagogiques en général.

L'alignement curriculaire élargi apparaît ici comme une lunette analytique compréhensive des contenus à enseigner, offrant la possibilité de mieux les organiser, les hiérarchiser et les faire circuler dans les évaluations. Nous observons toutefois que ses composantes semblent ne pas toutes avoir la même influence sur les réflexions et les pratiques de Fernand. C'est la raison pour laquelle, dans la discussion conclusive de notre manuscrit, nous tenterons de voir, à la lumière de tous nos portraits, quelles composantes de l'alignement curriculaire élargi permettent aux enseignants d'approfondir leurs analyses des ruptures au fil du temps.

Tout au long de la recherche

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse que Fernand est davantage outillé pour éviter des désalignements entre les objectifs, les tâches et les systèmes de pondération mis en œuvre, même s'il reste toutefois en difficulté au niveau de la construction de la note.

Nous appuyons nos hypothèses par trois mouvements en marche dans les pratiques évaluatives sommatives de Fernand, que nous considérons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi :

- Le premier articule la pondération et les objets, et relève d'une autre manière d'attribuer les points : « *l'attribution des points dans le travail, dans ce qui est évalué, dans les*

exercices. Avant c'était par gestes, maintenant, c'est vraiment lié aux progressions d'apprentissage » (entretien 2, FP 48) ;

- Le deuxième est lié aux contenus évalués et à leur proportion : *« je peux mieux évaluer l'équilibre du test, des différentes parties » (entretien 2, FP 48). « Je trouve que j'arrive mieux, en tout cas, j'ai une conscience et je fais des choix. Je vois, je dis « ah ben tiens entre ce qui est de ce thème et ce qui est le précédent, c'est à moitié-moitié ? Non, c'est trop. Il faut que je diminue ce qui a été appris précédemment pour que pour ceux qui l'ont pas, ce soit, je veux dire, que ça cache pas le travail qu'ils ont fait maintenant ». Donc ça me permet de faire des choix en conscience » (entretien 2, FP 53). Les contenus sont davantage analysés, mis en rapport avec l'enseignement et l'apprentissage. Le but : améliorer la visibilité du travail de l'élève. Là, Fernand dit : « Avant, je me disais « ah, ben c'est ça que je vais leur faire faire » et puis j'essayais de les faire travailler là autour. Là, j'ai l'impression que vraiment, je suis plus pointu. J'ai gagné en finesse ». (entretien 2, FP 63).*
- Cet élément induit le troisième changement de pratique, relatif à la cohérence des épreuves : *« quand je mettais la note et que j'avais, ce comme je vous avais déjà dit, ces impressions de « ah lui il a 4, mais j'ai l'impression qu'il n'a rien compris, et lui il a 3.5 mais il a plus compris », mais... voilà. Mais ça, j'ai l'impression que ça a disparu ». (entretien 2, FP 48-49). Nous pouvons y lire une opérationnalisation du modèle.*

Fernand dit avoir pu, grâce au modèle de l'alignement curriculaire élargi, expliciter certains ressentis : *« je pense que j'avais ça en moi, sans le savoir, inconsciemment, donc ça a permis vraiment de mettre, comme je redis, des mots de... de définir les choses. Voilà, ça a permis aussi de prendre conscience parce que je m'étais pas posé la question en termes d'évaluation de la qualité ou de la quantité » (entretien 2, FP 129). Puis il détaille la manière avec laquelle cette notion d'alignement a raisonné chez lui : « Mais alignement dans le sens de cohérence [...]. C'est vraiment de cohérence, ça veut dire que l'alignement, c'est moi qui le choisis, je choisis la ligne, et après, j'essaie de poser les éléments dessus. Et pas le fait de venir me mettre sur une ligne » (entretien 2, FP 130-135). Autant de réflexions qui sont visibles dans les deux épreuves réalisées.*

Mais qu'en-est-il maintenant du point de vue du développement potentiel des compétences évaluatives de Fernand ?

2.2 Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Fernand

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Fernand au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe. Pour répondre à cette question, nous mettrons en correspondance les apports déjà dégagés à l'aide des catégories conceptualisantes avec les compétences construites pour notre étude dont le développement est présenté dans la partie méthodologique.

2.2.1 Les compétences évaluatives développées principalement par Fernand

Il nous semble ici important de reprendre chaque compétence et d'en justifier le possible développement. Par souci d'explicitation, nous les rappelons ici¹³³ :

- A. Analyser le plan d'études du point de vue des objectifs et des objets de savoir disciplinaires qui y circulent à des fins d'évaluation sommative en mettant en relation de manière cohérente les différents niveaux de formalisation des objectifs, et en analysant les objets de savoir pour en déterminer les caractéristiques essentielles à des fins d'évaluation sommative.
- B. Elaborer des tâches évaluatives complexes en rapport avec les objectifs des plans d'études qui donnent des preuves d'apprentissages maîtrisés ou en voie de réalisation, en privilégiant des tâches qui donnent à voir des unités d'apprentissage significatives (approche qualitative).
- C. Adopter des systèmes de pondération permettant de mettre en évidence le poids donné à l'apprentissage référé aux objectifs et aux objets de savoir en privilégiant, tant que possible, une pondération critériée, notamment pour permettre un feedback aux élèves à visée de régulation.
- D. Construire des notes de manière à ce qu'elles rendent compte des acquis et lacunes des apprentissages des élèves en élaborant ses propres échelles et barèmes en référence au système de pondération, à la complexité des tâches et aux objectifs évalués.

¹³³ Nous ne répéterons toutefois pas ces compétences dans chaque portrait, par souci d'alléger le propos.

- E. Penser ses pratiques évaluatives sommatives en termes de validité et de cohérence d'alignement curriculaire élargi en les référant aux processus d'enseignement et d'apprentissage, et en analysant leur cohérence à l'aune d'éléments taxonomiques.
- F. Justifier ses choix en matière de pratiques évaluatives sommatives au regard de son contexte institutionnel cantonal, local (établissement) et de la culture des équipes dans lesquelles elles s'inscrivent, en les adaptant aux prescriptions institutionnelles, tout en étant attentif à ce qu'elles soutiennent les apprentissages des élèves.

Aussi, en nous appuyant sur nos résultats précédents, nous pouvons documenter quelles compétences Fernand¹³⁴ semble avoir développées :

Tableau 7 : Indices de développement potentiel des compétences évaluatives de Fernand

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Oui, référé aux contenus	Oui, en rapport avec les objectifs dans une perspective analytique			
B	Oui, traduisent des objets clés de la discipline présents dans le PER	Oui, traduisent des objets clés de la discipline	Oui, toujours complexes du point de vue des habiletés cognitives et des contenus		
C	Oui, les points sont donnés en référence aux contenus issus du PER	Oui, les points sont traduits depuis des critères portant sur des objets		Plus ou moins, les critères étant systématiquement convertis en points	
D	Oui, la note doit avoir un rapport avec les contenus issus du PER	Oui, la note doit rendre compte d'éléments de contenus	Oui, la note traduit la complexité des objets présents dans les tâches	Oui, la note est issue d'un raisonnement référé aux unités d'apprentissages pour le calcul du seuil du 4	Non, la note est toujours globalement compensatoire, ruptures probables
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures probables
F	Oui, rapport distancé	Oui, choix assumés	Oui, choix assumés	Non, le système critérié étant considéré comme non mutualisable	Non

¹³⁴ C'est en triangulant nos données et en observant la récurrence de certains propos ou pratiques que nous avons validé, ou non, le développement des compétences au travers de leurs multiples dimensions.

Nous l'avons montré dans la partie conceptuelle présentant l'AL, la perspective que nous en adoptons se traduit au travers de cinq caractéristiques, que nous reprenons pour montrer plus finement le développement des compétences évaluatives de Fernand. Pour la clarté du texte, nous les récrivons en italique dans la section suivante. Il nous semble intéressant de mettre ces différentes caractéristiques en dialogue avec nos résultats pour donner à voir le processus de développement des compétences de Fernand.

2.2.2 Le processus de développement des compétences évaluatives chez Fernand au regard des caractéristiques de l'*assessment literacy*

1. *La prise en compte du contexte dans lequel s'inscrivent ses pratiques.* Le questionnement de départ de Fernand sur la pondération est resté tout au long du processus de formation. Il a pu évoluer au regard de son contexte institutionnel, de ses populations d'élèves, mais aussi de ses collègues directs et des contraintes prescriptives qu'il a souvent mentionnées. Toutefois, c'est ce même contexte qui semble être un frein au développement de pratiques de notation référées à l'apprentissage.
2. *La négociation par Fernand des objets disciplinaires en jeu dans l'enseignement, l'apprentissage et l'évaluation.* Fernand choisit les objets disciplinaires qu'il va traiter, et surtout la manière avec laquelle il va les enseigner, en fonction de plusieurs facteurs pensés de plus en plus en référence à l'alignement curriculaire élargi. Ce processus décisionnel, qui caractérise l'AL, nous fait penser à une action de négociation que Fernand mène à chaque fois qu'il s'agit de savoir ce que les élèves vont devoir apprendre.
3. *Sa prise en compte des processus d'enseignement et d'apprentissage pour penser les pratiques évaluatives.* Dès le départ, Fernand pense ses pratiques évaluatives au regard des processus d'enseignement et d'apprentissage. Ce qui évolue, c'est ce sur quoi porte sa réflexion. Alors qu'au début il est davantage conditionné par les tâches, c'est en référence aux objectifs et aux contenus qu'il se réfère progressivement pour organiser les activités en classe puis dans l'évaluation sommative.
4. *Les connaissances incontournables en évaluation de Fernand référées à son contexte.* Fernand parle de la taxonomie comme une connaissance lui permettant d'être plus précis sur ses attentes, notamment au niveau de l'explicitation des apprentissages : « *Et on voit bien qu'appliquer une règle, en termes... dans la taxonomie, c'est une tâche bien plus complexe que de comprendre* » (entretien 2, FP 159). Nous faisons également l'hypothèse que le modèle de l'alignement curriculaire élargi lui a permis d'affiner sa

connaissance du PER, des caractéristiques des tâches évaluatives, et des systèmes de pondération, au vu de ses propos mais surtout des épreuves créées. En revanche, ses connaissances en notation - comme celles relatives à la nature des échelles - semblent encore en construction, notamment parce qu'elles se heurtent à des contraintes contextuelles qui les fragilisent.

5. *Une pratique évaluative considérée comme relevant, pour Fernand, d'un compromis permanent entre différentes dimensions.* Ces compromis sont très explicites lorsque Fernand exprime ses dilemmes à mettre en œuvre une évaluation critériée et une notation uniquement basée sur des critères. Nous observons concrètement ces compromis dans les deux systèmes de pondération et de notation qu'il élabore dans ses deux épreuves, lorsqu'il est prêt à construire un seuil de suffisance en fonction des apprentissages, mais pas pour les autres notes. D'autres compromis sont également présents dans son rapport au PER, lorsqu'il reformule ou sélectionne des objectifs, ou quand il décide de l'importance des objectifs à évaluer.

2.3 En guise de synthèse provisoire

Au regard de ce qui précède, nous observons que les compétences A, B et C semblent se développer significativement, même si au niveau de la dernière, l'opérationnalisation d'une évaluation critériée répondant à un souci de cohérence d'alignement curriculaire élargi est encore et toujours source de tensions.

C'est au niveau des compétences D, E et F que des résistances empêchent, dans une certaine mesure, leur développement. Dans la D, la persistance d'une construction de note dans laquelle différentes parties de l'épreuve ou dimensions de l'apprentissage évalué se compensent peut potentiellement induire des ruptures d'alignement curriculaire élargi. Dans la E, c'est cette même note qui peine à rendre compte complètement des apprentissages des élèves, acquis en termes d'habiletés cognitives et de contenu. Et au niveau de la F, c'est le contexte tel que perçu par Fernand (rapport entre les collègues, prescriptions légales) qui l'empêche en partie de partager et mettre en œuvre des pratiques novatrices de pondération et de notation.

Par ailleurs, en référence à notre acception de l'AL, nous observons qu'une tension présente sur une des catégories entraîne de fait une résistance au développement de plusieurs compétences. Il en va ainsi de la question de la pondération et de la notation, dont nous avons déjà parlé. Mais nous constatons également que des contraintes liées notamment aux objets

disciplinaires peuvent entrer en conflit avec le développement de ces mêmes compétences. Ainsi, par exemple, nous remarquons que des éléments stabilisés par Fernand dans sa première épreuve - comme le rapport cohérent existant entre les objectifs annoncés et les tâches évaluatives - se sont avérés plus problématiques dans la seconde. Cela confirme que les compétences évaluatives, dans cette perspective, ne sont jamais acquises une fois pour toutes ou se développent de manière non linéaire, car elles dépendent étroitement de multiples facteurs interagissant ensemble.

Fernand semble donc avoir acquis des connaissances et des compétences clés en s'appropriant le modèle de l'alignement curriculaire élargi. Nous pensons qu'une posture réflexive est en voie de développement grâce au processus de formation, au cœur de laquelle son jugement professionnel a gagné en cohérence grâce à la référence permanente à un modèle théorique¹³⁵. Certes, des impossibilités perdurent pour lui dans ses pratiques. Mais elles peuvent être résolues au travers de compromis qui n'invalident en rien la cohérence de ses pratiques pédagogiques.

3 Portrait 2 : Paul, enseignant de mathématiques

La réalisation du portrait de Paul s'appuie sur les données suivantes :

- Le premier entretien et le premier écrit, réalisés tout deux avant la formation ;
- Sa première et seconde évaluation¹³⁶ ainsi que leur discussion proposée par le formateur, son deuxième écrit, tous produits durant la formation ;
- Les échanges entre Paul, ses pairs et le formateur durant les séminaires de formation 2 et 3.
- Le second entretien et le troisième écrit, réalisés à la fin de la formation.

Abordons maintenant les dimensions interrogées par nos questions et sous-questions de recherche.

¹³⁵ Comme nous l'annonçons dans notre partie 6 sur la méthodologie relativement à la recherche collaborative, nous assumons pleinement que la référence au modèle théorique nous intéresse dans une perspective de transformation des pratiques des enseignants.

¹³⁶ Paul a réalisé seul ses deux épreuves, ne collaborant avec aucun des deux collègues du même établissement.

- 3.1 Les dimensions des pratiques évaluatives de Paul sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.

Avant la formation

- 3.1.1 Les dimensions de la pratique évaluative sur lesquelles Paul porte son attention *avant la formation*

Tant au niveau de son premier écrit que de son premier entretien pré-formation, Paul commence par dresser une réflexion générale sur l'évaluation, réflexion qui porte en partie les raisons de sa participation à la recherche. Cette réflexion porte sur les finalités de l'évaluation, qui, telles que présentes dans son contexte, lui semblent peu convenir : *« j'ai l'impression qu'on utilise l'évaluation que pour orienter l'élève ; on l'utilise jamais, enfin, en tout cas, ce que je constate, c'est qu'on utilise rarement l'évaluation pour que ça serve à l'élève directement, pour qu'il se dise : voilà, je sais ça, il faut que je bosse plus telle ou telle chose »*. Dans son écrit, il va encore plus loin en affirmant : *« cet aspect mécanique de l'évaluation ne me plaît pas du tout »*. Un aspect mécanique incarné notamment selon lui dans l'exploitation systématique du barème fédéral, qui ne permettrait pas de noter correctement, et serait donc à la base d'une orientation scolaire parfois peu étayée.

C'est d'ailleurs le problème principal que Paul soulève dans son entretien : la question des barèmes et de la notation, qui lui pose des questions insolubles, et parfois, l'emmène dans des considérations que l'on pourrait comprendre comme contradictoires. En effet, il affirme d'un côté être favorable à une forme de standardisation des évaluations : *« enfin, ce que je trouverais intéressant, ce serait d'arriver à ce qu'on évalue tous la même chose, même si je vois que c'est pas possible, mais aller dans ce sens-là ; qu'un élève, qu'on prend ici, ait les mêmes résultats à Pully. [...] Du coup, il faudrait que ce soit la même personne qui le fasse. Je sais qu'on y arrivera pas, mais j'aimerais aller dans ce sens-là »*. Et de l'autre, il affirme être très soucieux du niveau de ses élèves et dit faire du bon travail en évaluation, puisqu'il ne modifie jamais les barèmes a posteriori et que globalement, il peut dire : *« j'ai l'impression que je fais bien mon job quoi ; on peut toujours faire mieux, mais j'ai l'impression de chercher, en tout cas, à évaluer ce qu'il faut, et pis à pas être trop vache, ouais, j'ai l'impression que ça va assez bien. Alors c'est vrai que cette histoire d'échelle ça me travaille à chaque fois, on se dit « ouais, mais... », parce que finalement, c'est facile, on n'a qu'à soit adapter l'échelle, soit mettre des*

points bonus. Donc c'est vrai que c'est pas... oui, il faut faire un choix, à chaque fois, il faut faire un choix sans regarder « ah bah si je change ça, y en a deux qui passent de 3,5 à 4 ou de 3,5 à 3 ». Ça c'est un peu... à chaque fois, c'est une réflexion ». Par ailleurs, nous observons que Paul juge notamment une bonne évaluation à l'aune de la moyenne de classe « si je devais choisir une moyenne de classe à atteindre, ce serait 4,5, parce que ça voudrait dire qu'y a assez d'exercices compliqués, qui vont les faire travailler, les faire réfléchir, peut-être les faire progresser un peu pendant le test, et pis ils ont quand même eu suffisamment de possibilités pour atteindre le 4 ». Il pratique les points bonus en fonction des résultats des élèves, et malgré tout, change parfois le seuil de suffisance de ses barèmes si les résultats sont insuffisants.

Ces éléments de tension autour de la notation sont en rapport étroit avec un questionnement ouvert chez Paul qui touche plusieurs autres dimensions de sa pratique évaluative sommative :

- La pondération, au travers d'interrogations relatives à une manière correcte d'attribuer des points ; au fait d'attribuer des points en fonction de la difficulté des tâches « *Je me dis, des fois, « ah, les problèmes sont... m'ont l'air assez difficiles, donc je fais une partie technique plus facile », ou alors, c'est des problèmes très faciles, alors je me dis « je fais une partie technique qui a peu de points, voire pas de partie technique – ça nous est arrivé une fois, je crois ».* Nous observons qu'il est également attentif au rôle que jouent les demi-points, est ouvert à une forme de compensation au sein de l'épreuve ;
- La construction des épreuves, qui obéit toujours à la création de deux parties, une technique et l'autre traitant de problèmes mathématiques, mais dont l'équilibre n'est pas clair pour Paul, notamment dans le rapport que ce *design* entretient avec la note : « *j'aime aussi assez ce côté de faire une partie accessible, donc la partie « technique » on va dire – plus facile – et une partie plus difficile. Je trouve que ça c'est pas mal, de... l'idée ce serait, pour moi, ce qui serait bien, ce serait d'avoir un 4 facile à avoir. Donc voilà, on a le 4 parce qu'on a compris la base, et pis dès qu'on passe au 5 et plus, c'est qu'on a vraiment bien compris. Et pis ça, je trouve que, quand je fais mes tests en général, j'ai l'impression, en tout cas, que ça reflète assez la réalité [...] être plus juste là-dedans. Je trouve que c'est pas facile ».*
- Le recours au PER comme référentiel d'objectifs « *Ça nous donne des objectifs, mais ça nous dit pas qu'est-ce qu'on doit évaluer. Parce qu'y a dix mille façons d'évaluer une compétence, différentes...* ». Ce qui semble toutefois provoquer chez lui une confusion entre certains niveaux du référentiel : « Ah je les [les attentes fondamentales]

reprends pas telles quelles, je les reformule. Ah alors... la progression, ben se retrouve dans ces attentes fondamentales, non ? Finalement, il me semble... en tout cas en partie ». Et ce qui lui fait dire que « la progression des apprentissages, c'est le contenu des exercices, peut-être ; et pis l'attente fondamentale, c'est l'objectif final à atteindre en bout de leçon », mais également regretter le fait de vouloir systématiquement trop évaluer.

Cela étant, divers éléments de sa pratique sont énoncés comme stables par Paul :

- Le fait que l'épreuve sommative est toujours produite quelques jours avant la passation et conclut un thème ;
- Son contenu est toujours référencé à la feuille de route¹³⁷ qui a organisé les objets enseignés et les situations d'apprentissage, fidèlement aux moyens d'enseignement.

Les deux sources de données montrent que Paul porte une attention particulière aux objets mathématiques et aux tâches qui sont susceptibles de les travailler et les évaluer. À ce titre, son écrit, qui s'apparente à une analyse de pratique, montre l'importance qu'il donne à ces éléments : *« Nous sommes dans le thème des fractions, un thème qui a du mal à passer car on touche à la représentation du nombre chez les élèves. La compréhension de l'objet « fraction », de son concept, est difficile. C'est intéressant de prendre ce thème car il est représentatif de l'apprentissage en mathématique, qui peut être composé de deux parties bien distinctes : le « comment ça marche », et le « qu'est-ce que ça représente ». Un élève peut très bien avoir compris « comment ça marche » les fractions, sans pour autant comprendre ce qu'elles représentent. Même dans des exercices de type situation problème, il est possible de s'en sortir en comprenant comment traduire la consigne en langage mathématique, sans en avoir saisi le fondement »*. Nous observons également cette attention lorsqu'il explicite le contenu des tâches et leurs possibles variations : *« une situation qu'ils ont jamais vue, mais qui fait intervenir des notions ou des procédures déjà rencontrées ailleurs, peut-être pas forcément ensemble non plus. Peut-être que pour arriver à la réponse, il faut utiliser une procédure, plus une autre, plus une réflexion...[...] la situation-problème peut être différente »*.

Globalement donc, Paul rend compte d'un questionnement ciblé sur la notation, qui résonne avec d'autres dimensions de sa pratique. Ce qui est intéressant ici, c'est à nos yeux de constater

¹³⁷ Ce terme, fréquemment utilisé, décrit un plan de travail sur lequel figurent les tâches d'apprentissage à réaliser, leur ordre, et parfois d'autres informations permettant à l'élève de travailler de manière autonome.

que des dilemmes, voire certaines contradictions sont présents dans ses propos. En effet, des normes extérieures (barèmes standardisés, moyennes de classe) semblent faire référence et cohabiter avec une attention portée sur une évaluation de l'apprentissage et des progrès des élèves ; le PER est réinterprété et côtoie des moyens d'enseignements et des exercices tirés de diverses sources « *soit j'invente, moi, des problèmes, soit je vais en chercher ailleurs dans des livres, sur Internet, et puis voilà* ». Par ailleurs, certaines pratiques, bien que discutées, ne semblent souffrir d'aucune remise en question profonde : la construction des évaluations en une partie technique et une autre de résolution de problèmes, leur pondération en pourcentages est relativement uniforme.

Pendant la formation

3.1.2 Les points de cohérence et de ruptures d'alignement observés par Paul dans ses pratiques d'évaluation sommative lorsqu'il se réfère au modèle de l'alignement curriculaire élargi *pendant la formation*

Rappelons ici brièvement les trois temps de la réponse à cette sous-question :

- Présentation des données brutes en nous centrant sur les parties qui ont suscité un débat ou qui mettent en évidence des zones de tensions, de discussions, et en respectant l'ordre chronologique dans lequel elles ont été produites (ici : épreuve 1, épreuve 2, écrit 2).
- Présentation d'un tableau de synthèse de ces éléments que nous interpréterons en référence aux catégories conceptualisantes que nous avons définies.
- Commentaires de nos résultats.

3.1.2.1 Première évaluation créée pendant la formation

Cette évaluation porte sur les fonctions et l'algèbre. Elle a été passée par des élèves de 10 VG niveau 2 (Voie Générale, niveau d'exigences supérieures). Elle est constituée de 5 exercices à côté desquels figurent les points, au total de 20. L'objectif du PER n'est pas explicitement posé, ni l'échelle.

La page de garde de l'épreuve est sommaire et a été réalisée par Paul après coup. Nous y trouvons le total des points, un espace pour la note et la consigne suivante : « Calculatrice autorisée, aide-mémoire non autorisé. Durée 45 min. Ecris bien tous tes calculs et réponds aux problèmes par des phrases ».

Les éléments de discussion relevés par le formateur concernent toute l'épreuve, à savoir :

- Une proposition d'objectif du PER est faite : MSN33 Résoudre des problèmes numériques et algébriques, en construisant, exerçant et utilisant des procédures de calcul, en modélisant une situation de proportionnalité, en explorant les propriétés de quelques fonctions ;
- En rapport avec cela, quelques éléments de progression des apprentissages du PER sont également suggérés, au niveau des éléments pour la résolution de problème ;
- À ce titre, le formateur propose la réflexion suivante : « Il serait instructif de voir quelles PA sont évaluées, à quelle hauteur, idem pour les composantes » ;
- Quelques questions en suspens clôturent la discussion : Toutes les caractéristiques essentielles des contenus sont-elles évaluées ici ? Quelle est la complexité des tâches proposées ? Sont-elles en correspondance avec les objectifs (l'exercice 2 par exemple) ? Comment réalisez-vous l'attribution des points ? Pourquoi établir une échelle de type « (nombre de points obtenus : 20) x 5 + 1 » ?

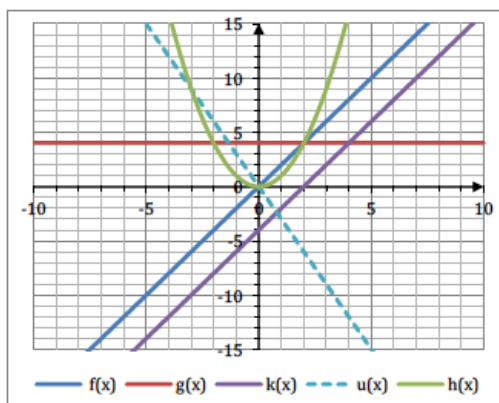
La discussion est donc volontairement amenée vers toutes les composantes de l'alignement curriculaire élargi, et porte potentiellement sur toutes les tâches, raison pour laquelle nous les reproduisons ici dans leur intégralité.

Exercice 1 (7 pts)

Voici 5 fonctions différentes et 5 tableaux de valeurs associés à chacune d'elles.

- Indique au sommet de chaque tableau le nom de la fonction correspondante (par exemple $g(x)$)
- Complète chaque tableau.
- Donne l'expression fonctionnelle des fonctions g et u :

$g(x) = \underline{\hspace{2cm}}$ $u(x) = \underline{\hspace{2cm}}$



x	y
2	
-2	4
4	
	4
	4

x	y
2	4
-2	-4
	0
4	
2	4

x	y
2	4
-2	4
	9
	9

x	y
2	
0	-4
4	
	10

x	y
2	
-2	6
4	
	10

Figure 19 : Paul, épreuve 1, fonctions et algèbre, tâche 1

Exercice 2 (2 pts)

Parmi les fonctions de l'exercice 1, quelles fonctions sont :

- ☐ Affine(s) : _____
- ☐ Linéaire(s) : _____
- ☐ Quadratique(s) : _____
- ☐ Constante(s) : _____

Figure 20 : Paul, épreuve 1, fonctions et algèbre, tâche 2

Exercice 3 (3 pts)

Voici le plan d'un appartement rectangulaire de 6.88m. Quelle est l'échelle de ce plan et quelle est la taille réelle du lit?



Figure 21 : Paul, épreuve 1, fonctions et algèbre, tâche 3

Exercice 4 (4 pts)

En mai, Monsieur Yamamoto achète un scooter pour 6 000 frs. Il s'arrange pour payer cette somme en trois fois. Il paye 40% du total directement au vendeur. En juin, il verse le 60% de la somme restante et en juillet il verse le reste. Calcule le montant exacte des trois versements.

Figure 22 : Paul, épreuve 1, fonctions et algèbre, tâche 4

Exercice 5 (4 pts)

Les remontées mécaniques de Veysonnaz voudraient construire un téléphérique qui amène les touristes de l'aéroport de Sion (altitude 481 m) à la station de ski de Veysonnaz (altitude 1011 m). Le trajet de ce téléphérique est représenté en rouge sur la carte ci-dessous, dont l'échelle est au 1 :30'000.

Quelle serait la pente moyenne de ce téléphérique, en % ?

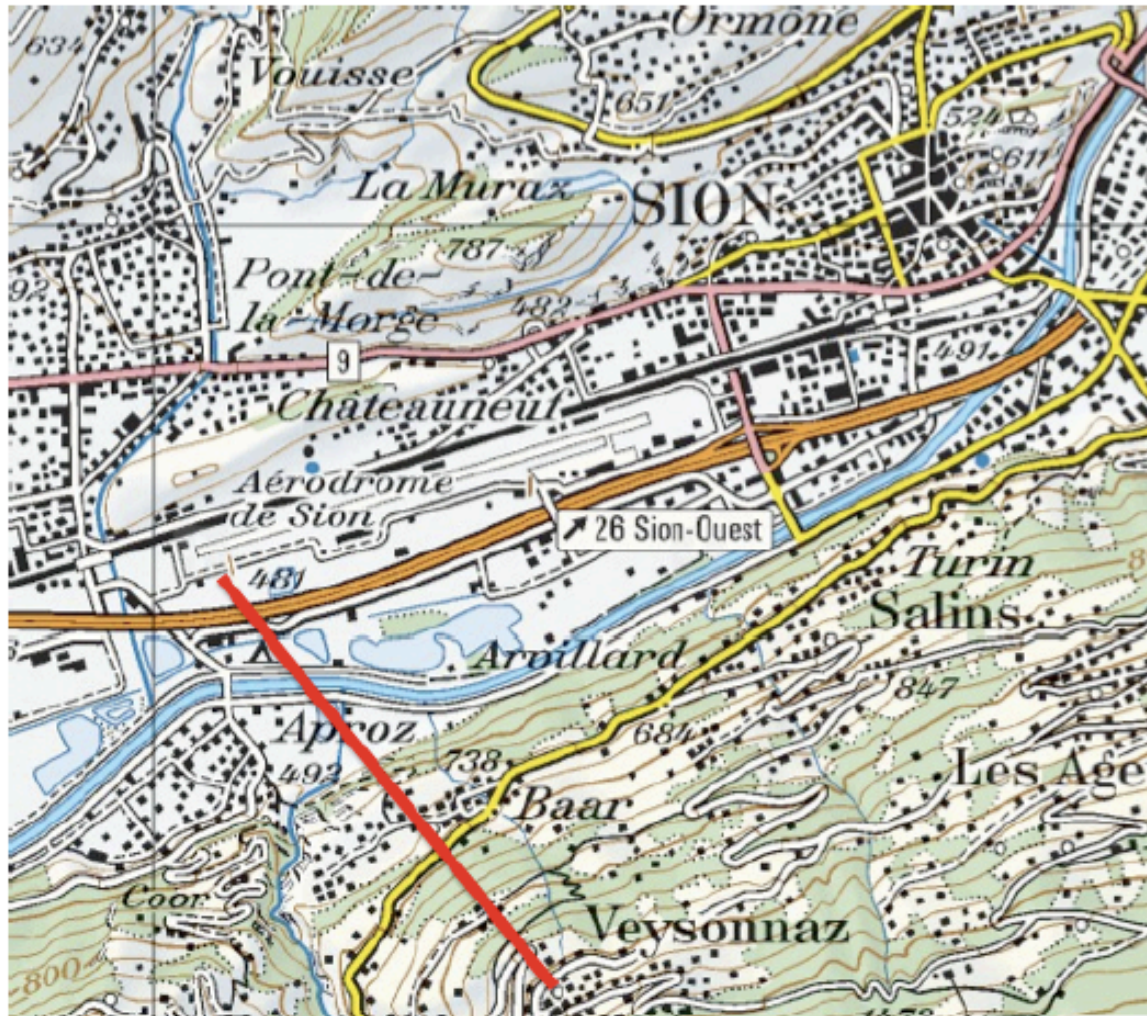


Figure 23 : Paul, épreuve 1, fonctions et algèbre, tâche 5

Lors du séminaire 2, Paul explique : « c'est un test que j'estime facile, et court. D'habitude j'évalue plus ». Il ajoute encore que c'est « un peu sur les fonctions, un peu sur la proportionnalité [...]. On commence à parler de la fonction comme objet mathématique ». « Il y a plein de petites choses, chaque problème c'est une de ces petites choses ». L'évaluation se réfère à une feuille de route, comme à son habitude.

3.1.2.2 Seconde évaluation créée pendant la formation

Six semaines se sont écoulées entre la discussion de la première et de la seconde évaluation. Cette dernière porte sur le chapitre « fonctions et algèbre : calcul littéral ». Elle se destine de nouveau à des élèves de 10 VG2, comme la précédente. Sur un document introductif, l'objectif est cette fois-ci mentionné (MSN33 Résoudre des problèmes numériques et algébriques), ainsi que ses composantes (au nombre de quatre) et des éléments de progression des apprentissages du PER tous pondérés de manière égale (au nombre de six). Il est intéressant de noter que Paul a mis en rapport chaque exercice avec la ou les progressions des apprentissages du PER.

Objectif du plan d'étude évalué : MSN 33 – Résoudre des problèmes numériques et algébriques

... en reconnaissant les caractéristiques mathématiques d'une situation et en la traduisant en écriture numérique

... en utilisant des propriétés des opérations

... en mobilisant l'algèbre comme outil de calcul

... en construisant, en exerçant et en utilisant des procédures de calcul

Progressions des apprentissages évalués:

P1 – Connaissance et utilisation des règles et conventions usuelles d'écriture algébrique (dans tout le test et exercice 1)

P2 – Détermination de la valeur numérique d'une expression littérale en substituant des nombres aux variables (exercices 8 et 9)

P3 – Elaboration d'expressions littérales à partir de problèmes, figures géométriques ou expressions verbales (exercices 7, 8, 9)

P4 – interprétation d'expressions littérales et identification de celles qui sont équivalentes (exercices 4 et 5)

P5 – Connaissance de la terminologie, écriture réduite et ordonnée de monômes et polynômes (exercices 1, 2, 3, 6, 7)

P6 – Opérations sur les monômes et polynômes (+/-/x) (exercices 2, 3, 6, 7)

Figure 24 : Paul, épreuve 2, calcul littéral, document introductif

Les éléments de discussion proposés par le formateur sous formes de questions permettant un débat sont ici les suivants :

- Une interrogation portant sur les éléments importants des contenus : sont-ils tous évalués ici ?

- Pourquoi certaines progressions d'apprentissage du PER sont évaluées dans plusieurs exercices et d'autres dans seulement deux ?

Sur ce même document, Paul explicite ensuite sa pondération et sa notation, qui suivent deux logiques : une référée à l'apprentissage et permettant de voir les seuils de suffisance par élément de progression des apprentissages du PER, l'autre plus traditionnelle par l'exploitation de points. Ses deux modes de faire étant particulièrement intéressants, nous les explicitons ici en référence aux documents produits par Paul, et parallèlement avec les éléments de discussion qu'ils ont suscités chez le formateur (la compensation possible entre les éléments de progressions des apprentissages du PER).

Echelle, notation

Je vais évaluer chaque progression avec la même valeur (ou pondération). Chaque progression sera évaluée séparément en fonction des résultats obtenus dans les différents exercices (par exemple, la progression P2 sera évaluée sur la base des performances réalisées dans les exercices 8 et 9). Chaque progression sera évaluée par une appréciation « insuffisant » - « suffisant » - « complètement acquis ». Le niveau de suffisance sera défini le plus clairement possible pour chaque progression.

Pour obtenir la note globale du test, je vais attribuer des valeurs aux appréciations, et le total obtenu sera comparé à une échelle que je vais définir.

La page suivante présente la grille des appréciations à utiliser pour la correction. L'échelle y est indiquée.

Sur le corrigé, vous trouverez également une notation en points. C'est ma façon de procéder habituelle avec l'échelle fédérale, ou une échelle créée en fonction de la dispersion des résultats. Je les ai indiqués à titre de comparaison.

Figure 25 : Paul, épreuve 2, calcul littéral, pondération et notation

Puis, Paul explique concrètement comment il va évaluer chaque progression et construire sa note à l'aide du tableau suivant.

Voici le tableau à utiliser pour l'évaluation. Lorsque l'appréciation pour une progression dans un exercice est insuffisante, on indique le chiffre 1. Si l'appréciation est suffisante, on indique le chiffre 2. Si la progression est considérée comme largement atteinte, on entre la valeur 3.

	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6	Ex 7	Ex 8	Ex 9	total
P1										
P2										
P3										
P4										
P5										
P6										

On peut donc obtenir au total 18 points au maximum sur l'ensemble de l'évaluation, ce qui représente la note de 6. Si toutes les progressions sont suffisantes, on obtient 12 points, c'est le 4.

Sur cette base j'ai créé l'échelle suivante :

note	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
points		6	7	8	9	10	12	13	15	16	18

Figure 26 : Paul, épreuve 2, calcul littéral, pondération et notation (tableau)

L'épreuve est constituée d'une page de garde et de neuf exercices. Le formateur n'est pas intervenu au niveau de ces derniers, au-delà des questions qu'il a formulées précédemment. Il a volontairement insisté sur les éléments de pondération et de notation, étant les plus révélateurs de changements et de questionnements chez Paul. C'est la raison pour laquelle nous ne les reproduisons pas ici, mais invitons le lecteur à se référer aux annexes Partie 4, 1.1.4 pour consultation.

Un commentaire est toutefois posé sur la page de garde, qui reprend les objectifs évalués, notamment à des fins de feedback pour les élèves. Là, le formateur dit : « Très intéressant de poser un espace qualitatif à visée de régulation en référence aux PA ». Nous y reviendrons plus loin.

TS math VG n2

Calcul littéral

Objectif du plan d'étude évalué : MSN 33 – Résoudre des problèmes numériques et algébriques

... en reconnaissant les caractéristiques mathématiques d'une situation et en la traduisant en écriture numérique

... en utilisant des propriétés des opérations

... en mobilisant l'algèbre comme outil de calcul

... en construisant, en exerçant et en utilisant des procédures de calcul

	Appréciation / Remarques
Conventions d'écriture	
Déterminer la valeur d'une expression par substitution	
Elaborer des expressions littérales	
Repérer et interpréter des expressions équivalentes	
Terminologie, écriture réduite	
Opérations (+ ; - ; x)	
Note obtenue	

Figure 27 : Paul, épreuve 2, calcul littéral, page de garde

3.1.2.3 Deuxième écrit réalisé pendant la formation

Son texte tient sur un peu plus d'une page (annexe Partie 5, 5.5). Les thèmes que Paul y met en évidence sont les suivants : les feedbacks, le rapport au PER et la notation.

Concernant le retour qu'il peut faire aux élèves suite à une évaluation sommative, il dit : « *je me suis rendu compte que je n'exploitais pas suffisamment le potentiel de feed-back d'une évaluation. [...] Idéalement, il faudrait rendre le test personnellement à chaque élève et le*

reprendre avec lui. Bien entendu ce n'est pas possible. Je trouve donc qu'une page de garde comportant des remarques ciblées est une bonne façon de mettre en évidence les forces et les faiblesses de chaque élève. C'est une manière de faire qui demande très peu de travail supplémentaire et que j'ai adoptée. Je me demande d'ailleurs pourquoi je ne l'ai pas mise en pratique plus tôt ». Cette prise de conscience raisonne avec ce qu'il a réalisé pour sa seconde épreuve, comme nous l'avons décrit plus haut.

En ce qui concerne le rapport au PER, Paul en montre une autre acception : *« En retournant à l'origine, donc en retournant voir ce qui se trouve dans le PER, les progressions et les compétences, je pense arriver à mieux créer les différentes tâches de mon test. En effet, les progressions d'apprentissage offrent une structure claire pour les tâches à insérer dans le test »*. Ce ne sont donc plus les moyens qui font référence, mais le plan d'études. Pour justifier ses dires, il convoque le modèle théorique : *« L'alignement curriculaire, pour être respecté, doit être basé sur le PER, et pas sur des moyens conçus par des tiers »*.

Enfin, relativement à la notation, Paul exprime quelques tensions. D'un côté il est en mesure d'explicitier le chemin parcouru vers une notation plus en lien avec l'apprentissage : *« Concernant l'échelle, j'ai beaucoup réfléchi, surtout lors de l'élaboration du dernier test [...] J'ai imaginé évaluer les élèves en prenant les progressions comme des critères, les unes après les autres et sur un pied d'égalité. On peut imaginer y ajouter d'autres critères, et les pondérer »*. Mais d'un autre, sa crainte d'une trop grande subjectivité semble prendre le dessus : *« Mais comment justifier que le choix de tel ou tel critère est plus équitable qu'un autre ? Sur le fond, je suis convaincu par ce type d'évaluation, mais en pratique, ce genre d'évaluation demande un travail conséquent, surtout au niveau de la conception [...] De plus, c'est une pratique qui n'est pas habituelle et peut déstabiliser certains élèves (et leurs parents !) Et finalement, nous devons attribuer une note, et comment transformer nos critères en note, de la façon la plus « juste » possible, et le problème reste le même « quelle échelle choisir » ? »*. Des éléments contextuels semblent par ailleurs également lui poser problème : *« Je vois aussi deux obstacles principaux à ces expériences : le temps à disposition et le cadre légal (et l'historique de l'école elle-même, basée sur une notation des élèves de 1 à 6 !) »*. Mais cela ne l'empêche pas d'entrevoir des changements dans ses pratiques.

3.1.3 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Paul à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *pendant la formation*

Sont prises en compte ici trois des quatre sources de données particulièrement significatives durant la formation :

- Les éléments de discussion proposés par le formateur (annexe Partie 5, 9). Mentionnés plus haut, ils se réfèrent tous au modèle théorique de l'alignement curriculaire élargi. Ils sont reproduits dans le tableau ;
- Les éléments débattus en séminaires de formation 2 et 3, retranscrits depuis les enregistrements ;
- Le deuxième écrit documenté précédemment.

Le tableau ci-dessous rend compte de la triangulation de ces données référée à nos catégories conceptualisantes.

Tableau 8 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Paul durant la formation

Données Catégories	Evaluation 1 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Paul débattus en séminaire 2 de formation	Evaluation 2 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Paul débattus en séminaire 3 de formation	Ecrit 2
Rapport au référentiel	Expliciter les objectifs, composantes et progressions d'apprentissage du PER évalués serait bénéfique	Le PER englobe bcp de choses dans la proportionnalité. Comment les élémenter ?	Pourquoi certaines progressions d'apprentissage du PER sont évaluées dans plusieurs exercices et d'autres dans seulement deux ?	Quels liens entre la feuille de route et l'évaluation ? ce n'est pas clair	« les progressions d'apprentissage offrent une structure claire pour les tâches à insérer dans le test »
		« j'utilise les composantes des objectifs car ils donnent les finalités de l'apprentissage, contrairement aux progressions du PER qui découpent			

		<i>trop</i> »			
Expression de l'objet	Les caractéristiques des objets sont-elles toutes évaluées ?	De quel objet mathématique parle-t-on ici ? De proportionnalité. Est-ce pareil dans tous les exercices ? Il y a des fractions, de la généralisation	Les caractéristiques des objets sont-elles toutes évaluées ?		
Elaboration des tâches évaluatives	Quelle est la complexité des tâches proposées ? Sont-elles en correspondance avec les objectifs (l'exercice 2 par exemple) ?	« <i>L'exercice 5 est la seule tâche complexe</i> »			Les tâches sont en rapport avec des éléments de progression des apprentissages du PER
Pratiques de pondération	Comment réalisez-vous l'attribution des points ?			« <i>Comment valider insuffisant (1)-suffisant (2)-complètement acquis (3) ?</i> »	Les progressions servent de critères
				« <i>Mettre un chiffre à côté de chaque appréciation me dédouane de mon choix</i> »	
				L'idée serait de permettre un retour aux élèves sur ce qu'ils réussissent ou pas	
Pratiques de notation	Pourquoi une échelle de type « (nombre de points obtenus : $20 \times 5 + 1$ » ?	Remise en question du barème fédéral : « <i>je peine à changer ça, j'ai peur de perdre en confort</i> »	La compensation entre éléments de progressions des apprentissages du PER est-elle possible ?	« <i>Comment passer de insuffisant (1)-suffisant (2)-complètement acquis (3) par progression d'apprentissage du PER à la note ?</i> »	Tension entre une note référée à l'apprentissage et un contexte contraignant
		Deux notes ici ? Une sur les échelles et pentes, une autre sur les			

		fonctions ?			
		Une alternative serait de chercher ce qui est attendu dans chaque exercice pour déterminer le seuil du 4			

Cette analyse au travers des catégories conceptualisantes donne à voir sur quelles dimensions de sa pratique Paul, avec l'aide de ses pairs et du formateur, porte ses réflexions. Par exemple, nous observons que dans la première épreuve, aucune référence au PER n'est clairement réalisée, alors qu'elle est argumentée plus tard. À ce niveau, Paul semble notamment en mesure de :

- Justifier de ses choix en matière d'objectifs évalués, et plus spécifiquement en ce qui concerne les composantes et progressions d'apprentissages du PER retenus ;
- Expliciter la nature des objets mathématiques en jeu dans les tâches des épreuves et leur rapport avec le référentiel ;
- Décider quand il semble plus opportun de se référer aux composantes ou aux progressions d'apprentissage du PER, notamment pour « *ne pas trop découper* » le contenu ;
- Faire des liens entre les exercices et les progressions d'apprentissage du PER.

Nous voyons également que certaines tensions semblent persister, comme, justement, le statut qui est donné aux différentes parties du PER. En effet, Paul est encore ici parfois hésitant puisqu'il exploite plutôt les *progressions des apprentissages* pour penser la notation et élaborer ses tâches, alors qu'il affirme mieux comprendre les *composantes* puisqu'elles se rapprochent des finalités des apprentissages mathématiques. Toutefois, ce nouveau raisonnement lui permet désormais de référer chaque tâche à des unités d'apprentissage, alors qu'au départ, il tenait à des épreuves divisées en deux parties, l'une « technique » et l'autre en « problèmes ».

Son acuité sur les objets mathématiques, quant à elle, semble perdurer. Un débat profond a lieu lors du séminaire 2, lors duquel est discuté la nature des objets en circulation dans sa première évaluation. Là, Paul est en mesure d'explicitier les contenus de chaque exercice, et de justifier que, pour lui, ils tournent tous autour de la proportionnalité. Pourtant, il doit faire face à un collègue qui aurait dissocié échelles et fonctions, et aurait donc attribué deux notes. Mais Paul de démord pas : « *on est entrés par les fonctions, puis on a travaillé les graphiques, aussi parce que c'est comme ça sur la feuille de route et dans le PER* ».

Mais le plus grand mouvement dans les pratiques de Paul se situe peut-être au niveau de la pondération. En effet, la seconde évaluation et la discussion qu'elle provoque, ainsi que certaines réflexions présentes dans son écrit, laissent penser qu'un changement de taille est en marche. Alors qu'avant d'entrer en formation il reste assez flou sur ses pratiques à ce niveau, il entreprend rapidement un travail approfondi sur un système de pondération centré sur des unités d'apprentissage, donnant à voir ce que l'élève réussit ou non. Le tableau de sa seconde épreuve est à ce niveau explicite : plutôt que de compter des points alloués à des exercices, il juge de l'atteinte de progressions d'apprentissage du PER au travers des exercices. Certes, son système insuffisant (1) - suffisant (2) - complètement acquis (3) questionne et le déstabilise : *« oui, effectivement, donner 1 à insuffisant pose problème, je me suis trompé, mais c'est pas important, j'ai qu'à changer l'échelle »*. Le formateur lui demande alors ce qui l'empêcherait de passer à un système validé/non validé pour déterminer la tendance de l'atteinte des progressions d'apprentissage du PER, pour éviter les points. Paul, avec ses collègues, dit que *« ça pourrait passer »*, et paraît prendre conscience de l'intérêt de la démarche pour mettre en évidence ce sur quoi porte la réussite ou l'échec de l'élève, puisqu'elle *« l'oblige [Paul] à retourner dans les exercices »*.

Mais toute cette réflexion semble à un moment donné se heurter à l'écueil de la notation. C'est là que le raisonnement engagé bute d'abord sur l'habitude de Paul d'exploiter depuis longtemps le barème fédéral qui lui donne somme toute satisfaction. Là, sa remise en question dudit barème rencontrée en début de formation est moins forte. Ensuite, arrive la tâche ardue pour lui de convertir son système d'appréciation en une note. La réflexion, dense entre lui et ses collègues, n'arrive pas à dépasser la contrainte de l'échelle à onze positions (des notes pouvant aller de 1 à 6 avec demi-notes) : comment rendre compte de cette recherche de qualité avec autant de nuances ? La question restera ouverte et momentanément non résolue.

À la lumière de ces éléments, nous sommes en mesure de formuler quelques hypothèses interprétatives traduisant des mouvements dans les pratiques évaluatives de Paul, et qui nous permettent de voir dans quelle mesure il parvient à travailler leur cohérence du point de vue du modèle théorique :

- Le PER devient une référence pour déterminer le contenu des évaluations, mais également pour juger de la cohérence entre l'enseignement et l'évaluation ;
- L'analyse des objets de savoir peut s'incarner dans le plan d'études dans une perspective critique ;

- L'analyse des tâches du point de vue de leur complexité et des objets mathématiques qu'elles convoquent est plus aisée, même si elle reste parfois peu explicite ;
- Une réflexion et une opérationnalisation d'un système de pondération complexe référé à des unités d'apprentissage est tentée, avec d'importantes limites toutefois.

Conjointement, il se dégage également des tensions ou difficultés à dépasser des ruptures d'alignement curriculaire élargi :

- Le rapport entre les éléments de progression des apprentissages du PER et leur composante n'est pas toujours explicite dans leur exploitation ;
- Bien que référée à l'apprentissage, la construction de la note ne peut être opérationnalisée, faute d'outils permettant sa transmission et à cause des échelles préconisées par le législateur.

Après la formation

3.1.4 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Paul à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi à la fin de la formation

Paul a produit son troisième écrit plus de trois semaines après le dernier séminaire de formation. Il l'a rédigé sous forme de bilan portant sur trois aspects faisant sens pour lui : ses attentes, la conception de l'évaluation, les échelles. Le tableau ci-dessous rend compte des éléments significatifs de deux sources de données relativement aux points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre : le dernier écrit et le second entretien. Ils sont directement mis en rapport avec les catégories conceptualisantes, de manière à rester synthétiques et centrés sur nos questions de recherche. Certes, des redondances peuvent apparaître, mais elles doivent à nos yeux être comprises et interprétées comme des signes d'un processus complexe d'appropriation de divers éléments par Paul. Nous commentons ces données plus bas en les mettant en dialogue avec les autres données.

Tableau 9 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Paul à la fin de la formation

Données Catégories	Ecrit 3	Entretien 2
	Rapport au « En math je peux m'appuyer directement	Retourner dans le PER pour « décortiquer

référentiel	<i>sur les moyens d'enseignement puisque ceux-ci sont conçus sur la base du PER. Mais du coup, je perds une part importante d'information, puisque j'utilise en quelque sorte des « compétences pré-mâchées »</i>	<i>l'apprentissage »</i>
	<i>« je me rends compte que si je ne base pas mon évaluation directement sur le PER, j'utilise mon vécu en classe et du coup certains objectifs sont parfois surévalués alors que d'autres sont négligés dans mon évaluation »</i>	Le PER est une nouvelle référence
		Un feedback possible sur les progressions d'apprentissage du PER
		Le PER est présent dans la feuille de route
		Le PER comme aide à l'équilibre des évaluations
Expression de l'objet		Des tests davantage référencés aux objets mathématiques, plutôt qu'aux tâches
		Centration sur ce que l'élève a appris
		Moins de contenus évalués
Elaboration des tâches évaluatives	<i>« Lors du choix des items à évaluer et des tâches correspondantes, je pense avoir trouvé un bon équilibre entre le respect du PER et mon enseignement en classe »</i>	<i>« plutôt un, une tâche, un exercice par objectif, et du coup, si un élève est, n'a pas saisi... un concept ou quelque chose qui fait partie d'un seul objectif, il est moins pénalisé »</i>
		Des tâches qui collent à un objectif, pas d'exercices à tiroir.
		Recherche de cohérence contenu-tâche-points
		Un seuil par tâche : intéressant, mais chronophage et pas possible pour tous les tests de mathématique
		Une autre manière de choisir les tâches
		Pas de cassure entre objectif et tâche
		<i>« Je mets des fois des questions bonus, mais c'est des questions qui sont peut-être un peu plus difficiles que les autres »</i>
		Des tâches qui demandent de mobiliser plusieurs choses en même temps
Pratiques de pondération	Un seuil de réussite par tâche, puis chaque tâche qui correspond à plusieurs objectifs, et pondérer les résultats obtenus selon l'importance de l'objectif : <i>« Le problème que je vois est que je risque rapidement de tomber dans la subjectivité, ce qui est à mon sens difficilement défendable (face aux parents, à la direction, aux autres élèves etc.) »</i>	Une recherche d'équité en référence au contenu : <i>« Mais si tout d'un coup, ben bêtement « Formulez une expression française en écriture littérale », si ça c'est une progression, au lieu de me dire « ah ben il a fait ça juste, je mets un demi-point », je vais me dire « ah ben ça vaut la même chose que la résolution »</i>
		Attribution des points en fonction de l'apprentissage à maîtriser
	<i>« C'est aussi plus facile pour moi, afin de rester constant dans mes corrections, de pouvoir m'appuyer sur un certain nombre de points »</i>	<i>« J'ai une autre vision, je pense, sur la façon de donner les points »</i>
		Critérier les progressions d'apprentissage du PER et y mettre des appréciations, ça change rien
		Evaluer par critères : <i>« Mais je vois mal, en maths, comment le faire. Parce qu'il y a peut-être trop peu d'éléments. Dans une évaluation, je sais pas, y a peut-être cinq... ouais, cinq progressions évaluées... je sais pas comment, je pourrais pas faire cinq paliers, c'est clairement pas possible »</i>

Pratiques de notation	« La note ne constitue pas un retour suffisant »	« que je choisisse le 4 à septante pourcents des points ou que je choisisse le 4 à trois quarts des critères réussis, c'est un peu pareil pour moi, c'est moi qui choisis... la part de subjectivité, elle est toujours importante »
	« Ma façon de procéder reste toujours sensiblement la même, donner un certain nombre de points à chaque tâche de l'évaluation, et établir une échelle arbitraire »	Difficile d'assumer des seuils par critères pour établir le 4
	« Je retombe sur l'obligation de choisir une échelle arbitraire »	Plus de difficultés à appliquer le barème fédéral, même s'il reste une base critiquable
	« j'ai approfondi ma façon de fixer le 4, en dénombrant les points minimaux à atteindre dans chaque tâches »	Pas de changements sur la construction de la note
		« la notation, donc le fait de choisir une échelle, ben ça pose un... c'est un nœud, comme vous avez dit avant. Ça peut former une cassure entre tout le travail que je me suis donné de la peine à faire avant... pour transposer... pour transposer les savoirs, pour faire un choix de, d'objectifs cohérents, que tout ce travail peut simplement être anéanti (rire)... par le choix d'une mauvaise échelle. C'est ce que j'ai retenu principalement »
		Cassure au niveau de l'échelle
		Prise de conscience que la notation fait partie de l'alignement curriculaire
		Pas d'autre solution que le barème standardisé
		« je pense qu'en améliorant la qualité de mes évaluations, donc dans leur conception, ben je suis sûr aussi qu'il ne devrait pas y avoir ce problème-là [des notes peu cohérentes] »
		« une évaluation où y a une distribution serrée, je pense que c'est une bonne évaluation »

Ce tableau nous permet de formuler des hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi :

- Le PER est une référence non seulement pour situer les objets mathématiques, mais aussi pour guider l'enseignement et orienter la construction des évaluations ;
- Les épreuves évaluent moins de contenu, mais de manière plus ciblée, plus alignée curriculairement ;
- Une distanciation d'un modèle d'épreuve dans lequel se côtoient des exercices techniques et d'autres de résolution de problèmes est visible. Les tâches tendent vers une complexité, sont indépendantes, référées à des objectifs ;

- L'attribution des points, plus sécurisante, se fait désormais en référence aux objectifs et aux contenus.

S'en dégagent également des tensions, ou difficultés à dépasser des ruptures d'alignement curriculaire élargi, surtout au niveau de la pondération et de la notation :

- L'exploitation de critères semble pour Paul impossible, notamment par le fait qu'il ne voit pas comment les pondérer pour aller vers une note. Cela dit, il perçoit tout de même du sens à tendre vers un jugement référé à des unités d'apprentissage pour donner du feedback aux élèves, mais sans aller jusqu'au verdict sommatif, comme dans sa seconde épreuve ;
- Faute de mieux, c'est donc le barème fédéral qui reste la référence pour construire les échelles chez Paul, mieux adapté à une pondération en points, et malgré le fait qu'il soit remis en question dès le début de la formation.

Nous observons donc ici deux mouvements à l'œuvre chez Paul dans ses pratiques évaluatives sommatives. Le premier dénote de profonds changements en cours que nous observons en triangulant nos données. Ces changements concernent une plus grande acuité des liens de cohérence qui devraient exister entre le plan d'études, les objets mathématiques à aborder via les différentes références à disposition (feuille de route, moyens d'enseignement, contraintes et habitudes du duo d'enseignants auquel Paul appartient), les processus d'enseignement et d'apprentissage et les tâches évaluatives. À ce niveau, les différences entre ses deux épreuves sont marquantes. Il est par ailleurs intéressant de noter que cette cohérence correspond à la compréhension que Paul a de l'alignement curriculaire : *« pour moi, l'alignement curriculaire, il s'arrête à la conception de l'évaluation. Après, quand on met la note, quand on choisit l'échelle, c'est plus dans l'alignement curriculaire »* (entretien 2, 184). Serait-ce une des raisons pour lesquelles le second mouvement, concentré sur les pratiques de pondération et de notation, aurait de la difficulté à s'opérationnaliser dans les pratiques ?

En effet, nous observons qu'à ces deux niveaux, de fortes tensions sont en jeu. Au niveau de la pondération d'abord, où critères et points semblent jouer le même rôle : *« signaler quand c'est pas réussi, et puis qu'est-ce qui est pas réussi. C'est ce que je pense, ce que j'aimerais faire. Alors... dans ma dernière évaluation que j'ai faite pour vous, là, j'étais, j'avais travaillé sur... plus des critères de réussite, avec des appréciations, et puis je trouvais ça pas mal. Mais je voyais pas comment passer de ces appréciations à la note. Enfin, je voyais difficilement. Y a,*

on peut toujours trouver des artifices, des trucs, mais ça revient toujours à choisir une échelle, à choisir un 4, un seuil de réussite, et puis après, ben voilà, de faire ça proportionnellement ou pas... c'est un peu toujours la même chose » (entretien 2, 3). Sa justification porte essentiellement sur le contexte : *« Même si j'aimerais passer sur des critères, mais je, je, c'est pas évident, ça demande trop de travail, et puis... Je, je, ouais, je suis pas prêt avec ça. Enfin je serais prêt, peut-être pour moi, mais je serais pas prêt à le justifier ensuite aux élèves, si peut-être aux élèves, mais aux parents, à la direction, peu... enfin je serais pas prêt à justifier, à dire que c'est cohérent, puis si on me dit « ah mais pourquoi t'as pas fait comme ça ? », j'aurais peut-être plus de peine. Tandis que si y a des points, ben je peux dire « c'est simple, y a un demi-point pour ça, un point pour ça, et voilà »* (entretien 2, 4). Mais elle met aussi en évidence, pour lui, une impossibilité de traduire un système de critères en une note.

C'est donc bien au niveau de la notation que perdurent les plus grandes impossibilités, ou, comme le dit l'intéressé, les compromis assumés : *« C'est ma base de travail [le barème fédéral]. Disons, parce qu'il faut bien travailler avec quelque chose. Je pense pas que c'est bien, mais il faut bien que je travaille avec quelque chose »* (entretien 2, 102-103). À ce niveau donc, Paul n'a rien changé dans ses pratiques. Certes, il n'en est pas pleinement satisfait, est conscient que cette procédure entraîne de fait une rupture d'alignement curriculaire élargi, mais s'en accommode, tant une solution alternative lui paraît improbable, voire risquée. Ce compromis est par ailleurs d'autant plus acceptable si la distribution des résultats reste régulière. Le barème fédéral, objet de questionnement dès le départ, prend donc ici le statut de référence.

Malgré ces derniers constats, nous faisons l'hypothèse que l'alignement curriculaire élargi apparaît ici comme une lunette analytique compréhensive des contenus à enseigner, permettant de mieux les organiser, les hiérarchiser et les faire circuler dans les évaluations. Nous observons toutefois que ses composantes semblent ne pas toutes avoir la même influence sur les réflexions et les pratiques de Paul. Si elles l'aident à penser ses pratiques jusqu'à, dans une certaine mesure, la pondération en points, elles peinent à l'aider à opérationnaliser une notation en rapport avec des unités d'apprentissage.

Tout au long de la recherche

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse que Paul est davantage outillé pour éviter des ruptures d'alignement - ou désalignements - entre les

objectifs, les tâches et les systèmes de pondération en points mis en œuvre. Cette hypothèse s'appuie sur l'observation suivante : de forts mouvements ont été opérés rapidement, entre les deux premiers séminaires de formation lors desquels Paul a été confronté au modèle théorique et aux premières réflexions sur la cohérence de l'évaluation sommative au travers d'échanges nourris avec ses pairs et le formateur.

Nous pouvons dès lors appuyer nos hypothèses interprétatives par trois changements importants en œuvre dans les pratiques évaluatives sommatives de Paul, que nous considérerons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi :

- Une attention est portée sur une évaluation plus juste, plus équitable, qui désormais porte davantage sur des contenus, poursuit le développement d'apprentissages complexes. Un des motifs de satisfaction de Paul à ce niveau est « *une évaluation qui me plaît en termes de contenu, qui est équilibrée, qui évalue les choses justes, mais qui soit pas trop difficile* » (entretien 2, 114) ;
- Le statut de l'objet mathématique dans cette recherche de cohérence est premier. En effet, nous retrouvons une centration sur le contenu lorsque Paul parle du plan d'études, mais aussi quand il explicite sa conception des tâches et leur rapport aux objectifs (voir notamment le travail effectué sur la seconde épreuve dans le document introductif), ou encore la manière avec laquelle il attribue les points : « *Avant, je faisais peut-être moins attention. Je faisais quand-même attention, hein, mais ça venait surtout lors de la correction. Lors de la correction, je voyais les différents cas, je voyais... les élèves qui avaient fait juste, des élèves qui avaient fait faux mais qui étaient partis dans une certaine direction qui me paraissait cohérente. Et lors de la correction, je me disais « ah, ben je pourrais accorder tant de points pour ça ». Maintenant, j'essaie d'anticiper, c'est pas... De me dire « voilà, si je leur donne ça, je peux regarder que cette partie soit juste ». Et quand je fais mes points, vu que je fais toujours des points, j'essaie d'être plus ou moins équitable* » (entretien 2, 42-43).
- Les tâches évaluatives sont pensées en référence aux apprentissages, même si leur complexité peine parfois à être explicitée.

D'un point de vue général, Paul dit que l'alignement curriculaire élargi l'a « *aidé à plus prendre conscience que la... la notation, donc le fait de choisir une échelle, ben ça pose un... c'est un nœud, comme vous avez dit avant. Ça peut former une cassure entre tout le travail que je me*

suis donné de la peine à faire avant...pour transposer...pour transposer les savoirs, pour faire un choix de, d'objectifs cohérents, que tout ce travail peut simplement être anéanti (rire)... par le choix d'une mauvaise échelle » (entretien 2, 178). L'analyse de nos données va dans ce sens. Il reconnaît donc être amené à « casser » cette cohérence au niveau de la notation par impossibilité de la penser différemment, mais il peut vivre avec, puisqu'il a bien amélioré sa pratique d'un point de vue général selon lui : « *Mais je pense qu'en améliorant la qualité de mes évaluations, donc dans leur conception, ben je suis sûr aussi qu'il ne devrait pas y avoir ce problème-là [de la notation]* » (entretien 2, 216).

Mais qu'en-est-il maintenant du point de vue du développement potentiel des compétences évaluatives de Paul, au regard de ces mouvements de pratique contrastés ?

3.2 Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Paul

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Paul au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe.

Il nous semble ici important de reprendre chacune des six compétences et d'en justifier le possible développement. Pour éviter de charger le texte, nous rappelons qu'elles sont décrites dans le portrait de Fernand qui ouvre ce chapitre. Aussi, en nous appuyant sur nos résultats précédents, nous pouvons documenter quelles compétences Paul¹³⁸ semble avoir développées :

Tableau 10 : Indices de développement potentiel des compétences évaluatives de Paul

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Oui, même si le statut des composantes et des progressions des apprentissage	Oui, en rapport avec les objectifs dans une perspective analytique			

¹³⁸ C'est en triangulant nos données et en observant la récurrence de certains propos ou pratiques que nous avons validé, ou non, le développement des compétences au travers de leurs multiples dimensions.

	reste parfois à clarifier				
B	Oui, traduisent des objets clés de la discipline présents dans le PER	Oui, traduisent des objets clés de la discipline	Plus ou moins, le statut des tâches complexes n'étant pas toujours explicite		
C	Oui, les points tendent à être donnés en référence aux contenus issus du PER	Oui, il y a une tentative d'attribuer des points depuis des objectifs portant sur des objets		Non, les points perdurent dans une perspective traditionnelle	
D	Oui, la note a un rapport avec le PER	Oui, la note rend compte d'éléments de contenus	Oui, la note tend à traduire la complexité des objets présents dans les tâches	Non, la note est issue d'un total de points. Seul le seuil du 4 est estimé en référence aux apprentissages	Non, la note est construite à l'aide d'un barème standardisé, compensations possibles, ruptures présentes
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures	Ruptures
F	Oui, rapport critique	Oui, choix assumés	Oui, choix assumés	Non, pas d'utilisation d'un système critérié, jugé impossible	Oui, les pratiques font l'objet de compromis jugés acceptables même si peu convaincants

Nous l'avons montré dans notre partie conceptuelle sur l'AL, la perspective théorique que nous en adoptons se traduit au travers de cinq caractéristiques, que nous reprenons pour montrer plus finement le développement des compétences évaluatives de Paul. Pour la clarté du texte, nous les récrivons en italique dans la section suivante. Il nous semble intéressant de mettre ces différentes caractéristiques en dialogue avec nos résultats pour donner à voir le processus de développement des compétences de Paul.

3.2.1 Le processus de développement des compétences évaluatives chez Paul au regard des caractéristiques de l'*assessment literacy*

1. *La prise en compte du contexte dans lequel s'inscrivent ses pratiques.* Si nous nous focalisons sur la pondération et la notation - deux dimensions de sa pratique sur lesquelles Paul a encore des doutes - nous constatons qu'une des raisons invoquées pour justifier d'un *statu quo* dans les pratiques est le contexte. Pour lui, c'est notamment parce que tout le monde utilise le barème fédéral qu'il sera difficile de s'en distancer.

Le contexte est donc ici parfois pris en compte pour développer des pratiques situées (s'adapter aux élèves), mais aussi expliciter des obstacles perçus comme insurmontables.

2. *La négociation par Paul des objets disciplinaires en jeu dans l'enseignement, l'apprentissage et l'évaluation.* Nous observons que le processus de négociation sur les objets disciplinaires s'inscrit désormais aussi dans une recherche de cohérence d'alignement curriculaire élargi. Des liens étroits sont réalisés entre les processus d'enseignement, d'apprentissage et les démarches d'évaluation.
3. *Sa prise en compte des processus d'enseignement et d'apprentissage pour penser les pratiques évaluatives.* Quand Paul affirme « Et en maths, comme on a ces moyens d'enseignement, je zappais peut-être le passage par le PER, en me disant « de toute façon, les moyens sont bien calqués sur le PER, donc... je basais mon évaluation sur les feuilles de route préparées par les collègues ou par moi, en fonction de ce qu'il y a dans les livres. Donc c'est quand-même basé sur le PER, c'est sûr puisque les livres sont basés sur le PER. Mais je pense quand-même qu'on loupe des choses. Et puis on... on focalise sur certaines tâches particulières alors qu'il n'y aurait pas lieu de le faire » (entretien 1, 1), il montre que l'enseignement, l'apprentissage et l'évaluation doivent être en cohérence du point de vue des apprentissages plus que des tâches.
4. *Les connaissances incontournables en évaluation de Paul référées à son contexte.* Nous avons déjà abondamment mis en évidence les composantes de l'alignement curriculaire élargi qui ont permis à Paul, comme autant de connaissances, de mieux appréhender ses pratiques évaluatives en contexte. Il n'en va pas de même de la taxonomie, jugée peu opérante : « je voyais mal comment utiliser cette taxonomie, à part pour juger les tâches, et dire « ah, ben ça c'est une tâche complexe, ça c'est une tâche un peu moins complexe... y a moins d'habiletés qui seront... utilisées... ». Mais au-delà de juste me dire à moi-même ah voilà, ça c'est une tâche complexe, je vois pas quelle est l'utilité dans le, dans le... dans la, l'élaboration de mes tests. Et pour les élèves, en fait. J'ai eu, et ça c'est vrai que j'ai de la peine... j'ai pas compris » (entretien 2, 221).
5. *Une pratique évaluative considérée comme relevant, pour Paul, d'un compromis permanent entre différentes dimensions.* Nous faisons ici l'hypothèse que les compromis de Paul se situent essentiellement dans la cohabitation, au sein de ses pratiques, d'une recherche de cohérence entre les objectifs et les tâches, avec une présence assumée de ruptures entre la pondération et la notation.

3.3 En guise de synthèse provisoire

Au regard de ce qui précède, nous voyons que c'est la compétence A, relative à l'appropriation du contenu du plan d'études et des objets de savoir, qui semble se développer principalement.

La compétence B, centrée sur la conception des tâches évaluatives, est également significativement en développement, même s'il demeure délicat de comprendre comment Paul justifie la complexité des tâches sans élément taxonomique.

En revanche, c'est au niveau des compétences C, D, E et F que des indices empêchant leur développement apparaissent. En effet, elles font toutes les quatre références à des dimensions relevant de la pondération et de la notation, celles-là même qui sont sources de tensions dans les pratiques de Paul. Il devient donc délicat de documenter, au regard de nos données, une progression à ce niveau, les questions et problèmes soulevés notamment au niveau de la transformation des points en note venant interférer avec une possible cohérence d'alignement curriculaire élargi.

Ici encore, en référence à notre perspective théorique d'AL, nous observons qu'une tension présente sur une des catégories entraîne de fait une résistance au développement de plusieurs compétences. Il en va ainsi de la question de la pondération et de la notation, dont nous avons déjà parlé. Ce constat revient à se demander s'il est justifiable de considérer qu'une évaluation curriculairement alignée entre les objectifs et les tâches, mais en rupture entre la pondération et la notation, peut être considérée comme cohérente et/ou valide ? Nous y reviendrons dans notre conclusion.

Paul semble donc avoir acquis des connaissances et des compétences clés en s'appropriant le modèle de l'alignement curriculaire. Dans son acception élargie ? Il est difficile de l'affirmer. Nous pensons toutefois qu'une posture réflexive est en voie de développement à certains niveaux grâce au processus de formation, au cœur de laquelle son jugement professionnel a gagné en cohérence grâce à la référence permanente au modèle. Sa tentative avortée de système de pondération critérié présent dans sa seconde évaluation en est selon nous un exemple pertinent. Sa situation est toutefois interpellante, d'un point de vue de formation. En effet, s'il prend conscience rapidement d'un certain nombre de phénomènes et semble s'approprier facilement certaines composantes de l'alignement curriculaire élargi pour penser ces derniers et modifier ses pratiques, d'autres semblent ne pas l'aider, voire même, paraissent ne pas du

tout lui parler. Nous serions même tentés de voir dans son parcours une position de repli opérée sur la fin au niveau de la pondération et de la notation. Des impossibilités perdurent donc dans ses pratiques, laissant chez lui un goût d'inachevé : « *En conclusion, j'ai pu trouver de bonnes idées et je pense avoir fait une grosse remise en question, bénéfique pour améliorer mes pratiques. Mais je ne suis pas prêt d'être satisfait à 100% de mes évaluations... Il y a encore du travail !* » (écrit 3).

4 Portrait 3 : Philémon, enseignant de mathématiques

La réalisation du portrait de Philémon s'appuie sur les données suivantes :

- Le premier entretien, réalisé avant la formation ;
- Sa première et seconde évaluation¹³⁹ ainsi que leur discussion proposée par le formateur, produites durant la formation ;
- Les échanges entre Philémon, ses pairs et le formateur durant les séminaires de formation 2 et 3.
- Le second entretien, réalisé à la fin de la formation¹⁴⁰.

Abordons maintenant les dimensions interrogées par nos questions et sous-questions de recherche.

4.1 Les dimensions des pratiques évaluatives de Philémon sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.

Avant la formation

4.1.1 Les dimensions de la pratique évaluative sur lesquelles Philémon porte son attention *avant la formation*

Philémon souhaite qu'au travers de cette formation, ses évaluations deviennent plus « *justes* » sur l'évaluation des compétences des élèves. Pour lui, l'évaluation devrait servir les apprentissages et il souhaite tendre vers cette finalité : « *Et puis dans l'idéal, effectivement, en*

¹³⁹ Philémon a réalisé ses deux épreuves avec sa collègue de classe parallèle Catherine, participant à la formation. Nous n'analyserons pas les données de cette dernière, vu qu'elle n'a pas écrit de récit et bien moins participé aux échanges en séminaires que Philémon.

¹⁴⁰ Philémon a décidé de ne pas écrire de récit de pratique. Nous établirons donc son portrait sur la base des autres données à disposition.

faire un outil, au moins pour quelques élèves, pour progresser, une fois qu'on rend l'évaluation, et puis pas en faire juste un passage arithmétique pour savoir si l'élève peut être promu, ou obtenir un certificat ou pas, mais en faire un outil pour l'apprentissage ». Plus concrètement, son questionnement principal porte sur deux dimensions :

- Une pondération qui ne permet pas toujours selon lui de mettre en évidence ce que l'élève maîtrise *« j'ai une manière de pondérer qui va se focaliser sur certaines choses et qui va être... qui va pas permettre toujours de faire émerger les compétences nouvellement acquises par un élève bien que travaillées. Et puis voilà, de ce côté-là, dire on sait, on a travaillé telle matière, est-ce que je suis vraiment en train d'évaluer chez l'élève cette acquisition de connaissances, ou est-ce que, ben y a l'arbre qui cache la forêt, pis en fait je suis en train d'évaluer tout le temps les mêmes choses chez l'élève »* ;
- Un statut donné aux prérequis qui n'est pas stabilisé *« dans les évaluations, surtout de tests significatifs, quelle part est donnée à la nouvelle matière, aux nouveaux acquis, et quelle part est donnée aux prérequis ? »*. Ce dernier questionnement réapparaîtra souvent, comme nous le verrons.

Ce qui frappe chez Philémon, c'est la quasi omniprésence d'une posture réflexive sur ses pratiques. Par exemple, lorsqu'il développe ce souci d'évaluer de manière *« juste »*, il fait allusion à ce décalage fréquent entre les notes attribuées et les acquis des élèves : *« est-ce que pour finir on peut avoir un élève qui a 5 de maths sur une échelle à 6 avec un seuil de suffisance à 4 – donc, je veux dire, c'est un élève à l'aise en mathématiques – alors que, si on prend ses évaluations, il arrive jamais à la bonne réponse... ? »*. Il en va de même lorsqu'il explicite ses pratiques évaluatives, organisées autour de quelques principes :

- La présence dans les épreuves d'exercices dits *« basiques »*, au nombre de cinq à six, pas davantage : *« Questions vraiment basiques, passages d'échelles, de l'un à l'autre, voilà, la distance sur la carte, la distance réelle, calculer l'échelle, avoir les trois situations. Dans ces situations-là, pas mettre trop de consignes, trop de français, même si c'est le tableau qu'ils ont déjà vu en classe, voir s'ils sont capables »* ;
- Une pratique *« récurrente »* visant à proposer ensuite une progression dans la complexité des tâches au sein des épreuves. Cela étant, dans une posture centrée sur les contenus, il se demande également comment évaluer des situations-problèmes dans sa discipline, tâche lui paraissant extrêmement délicate ;

- Un rapport étroit entre enseignement et évaluation : pas de nouveauté dans les contenus, aucun inédit, même s'il se demande « *si c'est juste ?* ».

Au travers de telles démarches, Philémon est soucieux de tout mettre en œuvre pour être sûr de voir ce que l'élève a acquis. Mais il n'en est pas certain : « *« est-ce que je suis vraiment en train de proposer des situations qui vont permettre d'évaluer ce que l'élève sait, ce qu'il est capable de faire ? ».* Donc voilà, c'est cette préoccupation, quand je parle de justesse... ». Il se questionne également sur ce principe d'évaluer des apprentissages de complexité différente : « *« est-ce que à un moment c'est juste d'être dans une évaluation, on va un peu saucissonner et on va essayer de mettre en exergue certains points, quitte à, voilà, à... Est-ce qu'on perd du bon sens, tout compte fait ? ».*

Il n'y a donc aucune certitude chez Philémon. Toutefois, ses propos sur son rapport au PER semblent plus ancrés, à quelques doutes près :

- Il le considère comme « *un fil rouge* » auquel il se réfère lorsqu'il souhaite savoir « *s'il est dedans* », mais reconnaît rapidement que sans cette référence, une évaluation devient un « *cul de sac* », puisqu'elle risque de s'arrêter à la note et non à l'apprentissage ;
- Dans ce sens, il s'autorise à le reformuler, car pour lui « *professionnellement, on doit s'autoriser à modifier les énoncés pour pouvoir les rendre intelligibles et compréhensibles au public qu'on a en face* » ;
- Toutefois - et cela pourrait sembler contradictoire - les objectifs qu'il poursuit et évalue sont davantage issus d'une « *routine* » et de « *l'expérience* » que du PER. C'est la raison pour laquelle il affirme que pour lui, les références sont les tests d'apprentissage ou les contenus nécessaires à l'insertion professionnelle¹⁴¹ : « *Donc, même si, à la limite, un plan d'études va me dire que c'est plus beau et plus ambitieux de travailler telle ou telle compétence, je veux que les élèves dont j'ai la charge puissent montrer de quoi ils sont capables en mathématiques dans un test d'entrée en apprentissage* ».

Relativement à la pondération, il dira que mettre des points et des demi-points aux exercices de base et aux plus complexes est « *quelque chose qui vient assez naturellement* », et sera capable de l'exemplifier en parlant de diverses situations, puisque pour lui, le couple pondération-apprentissage va de soi. Là, son questionnement semble disparaître derrière quelques

¹⁴¹ Rappelons que Philémon enseigne essentiellement en 11^e année, la dernière de la scolarité obligatoire, et qu'à ce titre, il doit préparer les élèves à l'entrée en apprentissage.

certitudes : pas de surpondération sur un exercice pour arriver à un total de points facile à distribuer, pas de sommes de points standardisées et une pondération en rapport avec les compétences. Cela ne l'empêche pas, parfois, de modifier la pondération a posteriori, même s'il avoue : « *j'aime pas trop, mais voilà...* ».

Quant à la notation, ses propos sont plus nuancés, mais portent sur de nombreux aspects eux encore en questionnement. Par exemple, il pense le seuil de suffisance (4) en référence à ce que l'élève doit savoir faire et des contenus importants : « *Par rapport au choix du barème... donc voilà, en mettant en place le corrigé, l'attribution des points, effectivement, je vais me dire « ben voilà, j'aimerais bien qu'ils soient capables de faire ça, ça et ça », je me dis « s'il est capable de faire ça, ça et ça, pis d'acquérir deux points dans la situation finale », je trouve que ça veut 4* ». Il y a déjà chez Philémon une réflexion autour d'une note référée à l'apprentissage qui distingue la maîtrise des outils et des compétences. Les contenus sont donc omniprésents dans ses réflexions, jusqu'à la notation, et un questionnement sur leur statut dans cette dernière phase est bien présent. Toutefois, sa réflexion sur le rôle des points et de leur rapport avec la note et le contenu semble ne pas être abouti : « *du moment que j'ai créé cette pondération, je vais pas aller rechercher ce qui, qu'est-ce qu'il y a derrière un point et demi là, et qu'est-ce qu'il y a derrière un point et demi là* ». Pourtant, il assume toujours poser le barème en amont puisqu'il est dépendant des contenus, et construit ensuite des échelles linéaires, même s'il dit être « *un peu pris au piège de cette linéarité...* ».

On le voit, Philémon traduit un nombre important de questions, voire de doutes sur ses pratiques évaluatives sommatives. Le fil rouge de ses propos tient dans une acuité sur les contenus, omniprésents, mais qui ne peuvent résoudre à eux seuls certains dilemmes. Les critères d'une « *bonne* » évaluation sont toutefois pour lui assez simples à cerner :

- « *Elle permet à l'élève d'être évalué sur ce qu'il a travaillé récemment, sur ce à quoi il s'est donné comme objectif, ou donné les moyens de maîtriser des compétences* » ;
- Elle va montrer à l'élève ce qui est réussi et échoué ;
- Elle met en évidence les prérequis non maîtrisés.

Pendant la formation

4.1.2 Les points de cohérence et de ruptures d'alignement observés par Philémon dans ses pratiques d'évaluation sommative lorsqu'il se réfère au modèle de l'alignement curriculaire élargi *pendant la formation*

Rappelons ici brièvement les trois temps retenus pour répondre à cette sous-question :

- Présentation des données brutes en nous centrant sur les parties qui ont suscité un débat ou qui mettent en évidence des zones de tensions, de discussions, et en respectant l'ordre chronologique dans lequel elles ont été produites (ici : épreuve 1, épreuve 2, écrit 2).
- Présentation d'un tableau de synthèse de ces éléments que nous interpréterons en référence aux catégories conceptualisantes que nous avons définies.
- Commentaires de nos résultats.

4.1.2.1 Première évaluation créée pendant la formation

Les données dont nous disposons ici sont particulières. En effet, Philémon a travaillé à l'élaboration de cette première épreuve en collaboration avec Catherine. Ils l'ont présentée lors du séminaire 1 et c'est bien cette version qui a fait l'objet de discussions. Mais suite à ce premier moment de formation, Philémon et sa collègue l'ont revue et en ont fait une seconde version qu'il nous semble ici également pertinente de convoquer, pour donner à voir concrètement les changements opérés et leurs justifications. C'est par ailleurs cette épreuve revue qui a été mise en œuvre en classe.

Cette évaluation porte sur les unités composées (vitesse et masse volumique). Elle a été passée par des élèves de 11e VSO¹⁴². Elle est constituée d'une page de garde sur laquelle figurent les 3 objectifs évalués avec leurs composantes (MSN 33-Résoudre des problèmes numériques et algébriques, MSN 34-Mobiliser la mesure pour comparer des grandeurs et MSN 35-Modéliser des phénomènes naturels, techniques ou des situations Mathématiques) ainsi que l'échelle adoptée. Un espace réservé à des feedbacks se trouve également sur le document.

¹⁴² Rappelons qu'au moment où se déroule notre recherche, la nouvelle Loi sur l'Enseignement Obligatoire (LEO) n'est pas encore entrée en vigueur en 11^e. A ce degré, c'est encore l'ancienne organisation scolaire qui fait foi, raison pour laquelle nous retrouvons encore des enseignements dans les trois voies (Baccalauréat, Générale et à Option).

Mathématiques : Test significatif
Unités composées (Vitesse et Masse Volumique)

Note :

Objectifs :	<p><u>MSN 33 — Résoudre des problèmes numériques et algébriques...</u></p> <ul style="list-style-type: none"> ▲ ...en reconnaissant les caractéristiques mathématiques d'une situation et en la traduisant en écritures numérique ou littérale ▲ ...en estimant un résultat et en exerçant un regard critique sur le résultat obtenu ▲ ...en modélisant une situation de proportionnalité <p><u>MSN 34 — Mobiliser la mesure pour comparer des grandeurs...</u></p> <ul style="list-style-type: none"> ▲ ...en connaissant le système international d'unités de mesures ▲ ...en calculant des grandeurs composées (vitesse, masse volumique,...) et en construisant les unités associées ▲ ...en mobilisant l'instrument et l'unité de mesure adaptés ▲ ...en exprimant une mesure dans différentes unités ▲ ...en calculant des grandeurs (aires, volume,...) <p><u>MSN 35 — Modéliser des phénomènes naturels, techniques ou des situations mathématiques...</u></p> <ul style="list-style-type: none"> ▲ ...en triant, organisant et interprétant des données ▲ ...en mobilisant des représentations graphiques (schémas, tableaux, graph,...)
--------------------	--

Echelle :

Points	26 pts	24 pts	22 pts	20 pts	17½ pts	16 pts	12 pts	8 pts	4 pts	0 pt
Note	6	5.5	5	4.5	4	3.5	3	2.5	2	1.5

Feed-back :

Tu connais... ou tu sais.....	Isolément	Dans un problème
... les formules		
... transformer les formules correctement		
... effectuer les changement d'unités		
... comprendre ce qui t'est demandé		
... calculer correctement avec la calculatrice		
... présenter clairement ton raisonnement, tes calculs		
... mettre en évidence tes réponses		

Mon commentaire :

Mon conseil :

Figure 28 : Philémon, épreuve 1, unités composées, page de garde

Des éléments de discussion sont déjà proposés ici par le formateur, à savoir :

- Des questions relatives aux objectifs posés : « Qu'est-ce qui vous empêche de prendre en considération les progressions d'apprentissage (MSN33 autour de la résolution de problèmes, MSN34 autour de l'estimation des grandeurs) ? Évaluez-vous vraiment tous ces éléments ? Et que dites-vous du fait que vous n'évaluez pas le MSN35 ? »
- Des questions autour du rapport entre le nombre d'objectifs et la notation : « S'il y a plusieurs objectifs évalués, pourquoi ne pas poser plusieurs notes ? Ou sont-ils complémentaires ? Inclus ? Quel est votre point de vue ? »
- Une interrogation autour des éléments de feedback, relevés comme intéressants, comme la recherche de cohérence du point de vue de l'alignement curriculaire élargi : « quels sont les éléments alignés ? ».

Du côté des contenus, cette évaluation est constituée de quatre exercices, certains subdivisés en sous-exercices, à côté desquels figurent les points, au total de 26.5. Pour une plus grande lisibilité, nous présentons les exercices en parallèle avec les éléments de discussion en rapport.

<u>TS vitesses et masses volumiques</u>				/ 26.5 pts
Tous les calculs qui te permettent d'arriver à la réponse doivent être notés clairement . Les réponses se donnent avec 2 chiffres après la virgule, arrondies correctement.				
1) Effectue les transformations suivantes:				/5pts
a) 134,5 m ³ = litres	1	d) 7428 s =hmins	1.5	
b) 7,89 t = kg	0.5	e) 675'000'000 cm ³ = m ³	0.5	
c) 0,15 h = min	1	f) 10 m/s = km/h	0.5	

Figure 29 : Philémon, épreuve 1, unités composées, tâche 1

Un élément de discussion est relevé par le formateur sur cette tâche :

- Une interrogation sur le fait que cet exercice, vu son faible niveau de complexité, pourrait avoir un statut de test assimilé (TA)¹⁴³ ;

L'exercice 2 - en non g (erreur de frappe sur l'épreuve) - s'apparente quant à lui à un problème.

g) Lors d'un cours de sciences, vous avez étudié les propriétés de quatre solides. Le tableau ci-dessous résume les résultats que vous avez obtenus. En quelle matière ont été construits ces solides ? (Aide-toi du tableau des masses volumiques ci-dessous.) /4pts

Forme du solide	Masse	Volume	Masse volumique (Utilise une unité usuelle)	Matière
Pyramide	2,3 kg	315.8 cm ³		
Sphère	222,5 kg	0.025 dm ³		
Cylindre	378 dg	14 cm ³		

Matière	Masse volumique en kg/dm ³
Aluminium	2,70
Fer	7,88
Etain	7,28
Argent	10,50
Or	19,30
Cuivre	8,90
Plomb	11,34

Figure 30 : Philémon, épreuve 1, unités composées, tâche 2

Les éléments de discussion proposés sont ici les suivants :

- Le fait que le recours au calcul ne soit pas obligatoire : « J'ai réussi à faire tout juste sans poser les calculs complètement corrects, par déduction. Qu'en dites-vous ? ;-) » ;
- Un questionnement sur le rapport entre la tâche et l'objectif : « Est-ce relatif au MSN34 ? » ;

¹⁴³ Il est possible de prendre en compte une série de travaux qui vérifient uniquement l'acquisition de connaissances ou de techniques spécifiques : les travaux assimilés (TA). Cet ensemble de travaux est assimilé à un travail significatif (TS) qui lui, ne doit porter que sur des apprentissages complexes (CGE, 2015, p.11).

- De nouveau, une interrogation sur le fait que cet exercice pourrait avoir un statut de travail assimilé (TA).

La partie 3 est constituée de quatre problèmes (a à d).

Effectue les problèmes suivants :

a) La distance entre Bex et Vionnaz est de 13,5 km. Je suis partie de Bex à 13h43 et je suis arrivée à 14h 28min 42s à Vionnaz. A quelle vitesse (en km/h) ai-je roulé ? /3pts

Figure 31 : Philémon, épreuve 1, unités composées, tâche 3a

b) Eliott parcourt un trajet de 7'500 m entre Gryon et Bex. Sachant qu'il roule à 18 km/h, à quelle heure arrive-t-il à Bex s'il part de Gryon à 10h53 et qu'il fait 6° ? /3pts

Figure 32 : Philémon, épreuve 1, unités composées, tâche 3b

c) Un avion a mis 59h 43min 40s pour faire le tour de la terre. Sa vitesse moyenne est de 690 km/h. Sachant qu'il a fait une escale de 4 heures, combien de km a-t-il parcouru ? /3pt

Figure 33 : Philémon, épreuve 1, unités composées, tâche 3c

d) Quel est le volume en cm^3 de 17 kg de nickel ? (Mv du nickel : 8.9 g/cm³) /2pts

Figure 34 : Philémon, épreuve 1, unités composées, tâche 3d

Là, le formateur intervient à deux niveaux pour engager la discussion de l'épreuve : le rapport entre les démarches de résolution et les points y relatifs, et le lien entre la complexité de chaque problème et leur pondération. Il dit : « Dans ces problèmes, ma façon de les résoudre, mes étapes, ne correspondent pas toujours aux points attribués. Cela m'interpelle... ».

L'épreuve se termine avec deux problèmes complexes, le 4a et le 4b.

- 4) a) Voici un récipient. L'intérieur de ce récipient est un cube de 28 cm de côté. Lorsqu'il est vide, il a une masse de 1,28 kg. On le remplit d'un liquide mystérieux. Le récipient a alors une masse de 14,82 kg, lorsqu'il est rempli. Quelle est la masse volumique (en kg/dm^3) mystérieux liquide ? /3pts

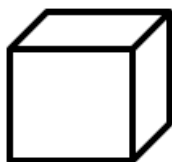


Figure 35 : Philémon, épreuve 1, unités composées, tâche 4a

- b) Voici un cylindre en argent (MV: $10,5 \text{ t/m}^3$), sachant que son diamètre mesure 72 dm et sa hauteur 112 cm, calcule son poids en kg ? /3.5pts



Figure 36 : Philémon, épreuve 1, unités composées, tâche 4b

Les derniers éléments de discussion sont les suivants :

- Une interrogation sur le rapport entre objectifs, tâches et attribution des points : « Avez-vous fait le rapport entre les apprentissages évalués, les tâches y relatives et leur pondération ? Est-ce possible ? Jusqu'où ? (l'exercice 1 vaut 5 points, alors que le problème 3c en vaut 3) » ;
- Une question sur les objets mathématiques en jeu dans l'évaluation : « Les caractéristiques fondamentales des contenus évalués ici sont-elles présentes ? ».
- Et une enfin sur la pondération : « Par ailleurs, je fais l'hypothèse qu'ils sont de complexité différente, mais ils valent à peu de choses près tous le même nombre de points. Pourquoi ? ».

Lors du séminaire 2, Philémon explique que cette évaluation est une adaptation d'une ancienne version. Trois éléments y ont notamment été ajoutés : les objectifs du PER, posés a posteriori ;

les points, attribués en fonction des contenus ; la grille de feedback en référence aux apprentissages poursuivis. Les échanges ont passablement bousculé Philémon. D'entente avec sa collègue, ils modifient leur épreuve avant de la faire passer en classe.

4.1.2.2 Première évaluation revue pendant la formation

Un mois après le séminaire 2, Philémon écrit le mail suivant au formateur :

« Rétroactivement, je pense que pour évaluer au plus juste les compétences de chaque élève, il aurait fallu combiner plusieurs compétences, plusieurs critères. Par conséquent l'attribution de points pour plusieurs critères (identification et expression de la formule nécessaire, conversion des unités utiles à la résolution, ...) me semble être un "pis-aller" plus juste quand on décide d'évaluer autant de compétences. En effet je n'arrive (pas encore) à sortir de cette forme de "pot-pourri" de compétences. Donc, la seule réponse que j'arrive à y donner, c'est un "pot-pourri" de points en tâchant d'agir le mieux possible sur la pondération afin que le pot-pourri se transforme en pot-au-feu savoureux. (Désolé pour ces métaphores hasardeuses, mais c'est tout ce qui me vient pour l'instant !!!) ». (correspondance personnelle, 22 avril 2015).

Dans ce message, produit spontanément, il discute et présente ses choix en matière de rapport entre les contenus et les points, qui représentent, pour lui, encore un dilemme. Pourtant, il effectue un changement important au cœur de cette épreuve : *« Pour revenir au TS modifié, nous avons convenu de raccourcir l'exercice 1 et là je me suis un peu emporté et je l'ai supprimé pour le convertir en un TA comme tu nous l'avais suggéré »* (correspondance personnelle, 22 avril 2015).

Dès lors, l'épreuve est divisée en deux parties : un TA sur les conversions d'unités à hauteur de 11 points aboutissant à une note, et un TS regroupant les quatre exercices de résolution de problèmes, à hauteur de 22 points, donnant lieu à une autre note. Ces documents n'ont pas été discutés directement, puisqu'ils ont été réalisés à la seule initiative de Philémon et de Catherine, suite aux échanges réalisés en formation. Ils peuvent être consultés en annexe Partie 4, 1.1.6.

4.1.2.3 Seconde évaluation créée pendant la formation

La seconde évaluation porte sur les transformations géométriques (symétries et homothéties). Elle se destine à des élèves de 11 VSO, comme la précédente. Elle est constituée d'une page de garde sur laquelle figurent l'objectif évalué et quatre éléments de ses progressions des

apprentissages issus du PER (MSN 31-Poser et résoudre des problèmes pour modéliser le plan et l'espace) ainsi que l'échelle adoptée. Un espace réservé à des feedbacks se trouve à nouveau sur le document.

Mathématiques : Test significatif Transformations géométriques (Symétries & Homothéties)										
Note :										
Objectifs :	<u>MSN 31 — points 4 et 7</u> <ul style="list-style-type: none"> ▲ À l'aide des instruments appropriés (règle, compas, équerre, ...) construction de l'image d'une figure plane par une isométrie : translation, symétrie axiale, rotation, symétrie centrale. ▲ Agrandissement et réduction de figures planes par une homothétie ▲ Anticipation de la position d'une figure plane après une ou plusieurs isométries ▲ Description et identification des caractéristiques d'une symétrie axiale et d'une homothétie 									
Echelle :										
Points	23	21	19	17	15	13	11	9	7	0
Note	6	5.5	5	4.5	4	3.5	3	2.5	2	1.5
Feed-back :										
Tu connais... ou tu sais.....										
-> effectuer une symétrie axiale :										
.....										
-> effectuer une translation :										
.....										
-> effectuer une rotation :										
.....										
.... effectuer une homothétie :										
.....										
As-tu nommer les points ? :										
As-tu travailler précisément ?										
As-tu travailler soigneusement ?										
Mon commentaire, mon conseil :										
.....										
.....										
.....										
.....										

Figure 37 : Philémon, épreuve 2, symétries et homothéties, page de garde

Là également, des éléments de discussion sont déjà proposés par le formateur, à savoir :

- Un doute quant au fait que la progression « Anticipation de la position d'une figure plane après une ou plusieurs isométries » soit évaluée dans un des exercices ;
- Une interrogation au niveau de la pondération qui met en valeur des aspects de la production indépendants de l'apprentissage visé : « Le soin, la propreté et la précision « pèsent » 5.5 pts, soit presque un quart du total, n'est-ce pas beaucoup ? D'autant que ce n'est pas un apprentissage que vous dites évaluer » ;
- Une réflexion mettant en évidence la contradiction potentielle existant entre le fait de proposer un feedback sur des éléments de l'apprentissage et une notation en points.

Cette évaluation est constituée de quatre exercices complexes, à côté desquels figurent les points, au total de 24. Pour une plus grande lisibilité, nous présentons les exercices en parallèle avec les éléments de discussion en rapport.

1. a) À partir de la figure FGH , dessine la figure $F'G'H'$ par une translation de vecteur v . /4.5 pts

b) Toujours à partir de la figure FGH , dessine la figure $F''G''H''$ par une rotation $(O ; -70^\circ)$. /4.5 pts

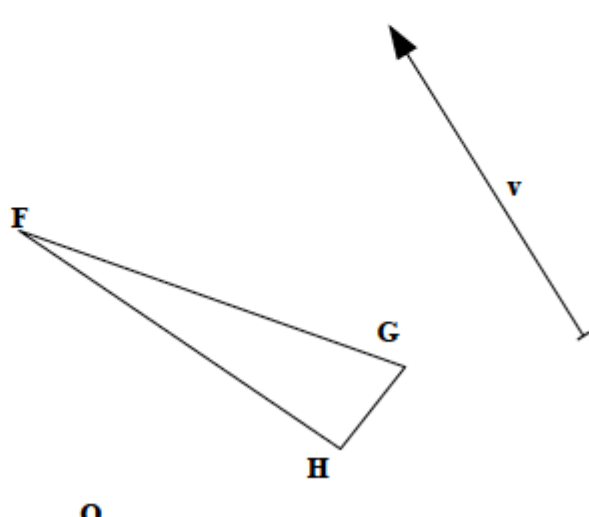


Figure 38 : Philémon, épreuve 2, symétries et homothéties, tâche 1

Globalement, les éléments de discussion portent sur deux aspects : l'attribution des points au regard des contenus et de la complexité des apprentissages évalués, et le rapport entre les éléments de progression des apprentissages du PER et les exercices. Par exemple, dans cet exercice 1, le formateur demande : « Globalement, je reste perplexe sur la plus-value des demi-points, et sur la façon dont vous les attribuez. Par exemple, dans cet exercice, F'' ne peut-il pas être considéré soit comme correct, soit comme erroné (1pt / 0 pt) ? ». Puis, il questionne si c'est bien la progression « À l'aide des instruments appropriés (règle, compas, équerre, ...) construction de l'image d'une figure plane par une isométrie : translation, symétrie axiale, rotation, symétrie centrale » qui est évaluée¹⁴⁴.

Le formateur propose le même travail réflexif pour les autres exercices.

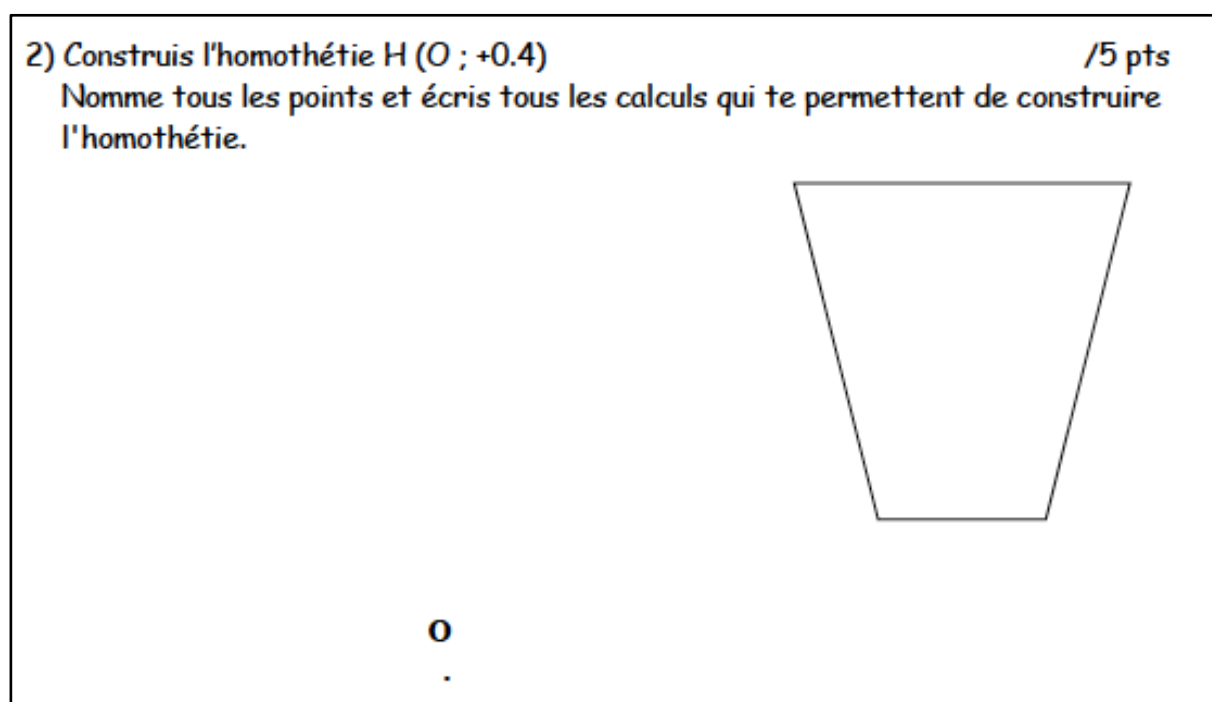


Figure 39 : Philémon, épreuve 2, symétries et homothéties, tâche 2

Là, il questionne si ce sont bien les progressions « Agrandissement et réduction de figures planes par une homothétie » et « Description et identification des caractéristiques d'une symétrie axiale et d'une homothétie » qui sont évaluées.

¹⁴⁴ Notons que pour des raisons de compréhension, nous encourageons le lecteur à se référer au corrigé de l'épreuve réalisé par Philémon présenté dans l'annexe Partie 4, 1.1.8.

3) La figure $ABCDE$ a pour image $A'B'C'D'E'$ dans une symétrie d'axe w .
 À toi de retrouver l'axe de symétrie et de redessiner les deux figures. /5 pts

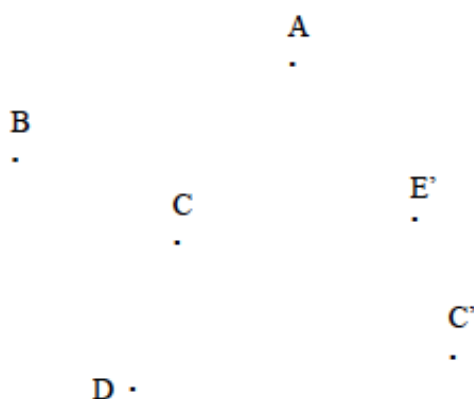


Figure 40 : Philémon, épreuve 2, symétries et homothéties, tâche 3

Là, si c'est bien la progression « À l'aide des instruments appropriés (règle, compas, équerre, ...) construction de l'image d'une figure plane par une isométrie : translation, symétrie axiale, rotation, symétrie centrale » qui est en jeu. Puis, dans le dernier exercice, il investit si ce sont bien les progressions « Agrandissement et réduction de figures planes par une homothétie » et « Description et identification des caractéristiques d'une symétrie axiale et d'une homothétie » qui sont évaluées.

4) Retrouve le centre d'homothétie, calcule son rapport (écris le calcul) et termine l'homothétie H de centre O

Rapport $r = \dots\dots$ (Arrondis le rapport à un chiffre après la virgule) /5 pts

Figure 41 : Philémon, épreuve 2, symétries et homothéties, tâche 4

On le voit, les éléments de discussion portent sur les mêmes caractéristiques des épreuves. C'est un élément intéressant que nous reprenons dans la suite de nos analyses.

4.1.3 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Philémon à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *pendant la formation*

Sont prises en compte ici trois des trois sources de données particulièrement significatives durant la formation :

- Les éléments de discussion proposés par le formateur (annexe Partie 5, 9). Mentionnés plus haut, ils se réfèrent tous au modèle théorique de l'alignement curriculaire élargi. Ils sont reproduits dans le tableau ;
- Les éléments débattus en séminaires de formation 2 et 3, retranscrits depuis les enregistrements.

Le tableau ci-dessous rend compte de la triangulation de ces données référée à nos catégories conceptualisantes.

Tableau 11 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Philémon durant la formation

Données Catégories	<i>Evaluation 1 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Philémon débattus en séminaire 2 de formation</i>	<i>Evaluation 2 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Philémon débattus en séminaire 3 de formation</i>
Rapport au référentiel	Statut des objectifs annoncés Rapport des objectifs avec les tâches Les caractéristiques fondamentales des contenus évalués ici sont-ils présents ?		Rapport des objectifs avec les tâches	La progression sur « l'anticipation du résultat » est bien évaluée
Expression de l'objet	Quel alignement entre les éléments de feedback et les objectifs ? Complexité des exercices et statut dans l'évaluation	Les objectifs « circulent » dans tous les exercices	Cohabitation d'un feedback relatif aux apprentissages avec une note en points ?	« je tiens au panachage de savoir-faire, je suis bien avec ça, avec cet aspect compensatoire » « j'arrive pas à privilégier tel ou tel savoir-faire, tout a de la valeur, mais pas la même »
Elaboration des tâches évaluatives	Interrogation sur le rapport entre	Les tâches vont du plus simple au plus		

	objectifs, tâches et attribution des points	complexe ¹⁴⁵		
Pratiques de pondération	La pondération correspond-elle à la complexité des exercices ?	La pondération peut permettre à un élève d'obtenir suffisamment de points sur une dimension de l'apprentissage pour ne pas être en échec, même si des lacunes persistent ailleurs	La pondération porte sur des aspects externes à l'apprentissage (propreté)	« je me suis pris au piège dans la mise des points en fonction de ce qui est attendu »
			La pondération correspond-elle à la complexité des exercices ?	Avec la propreté, évalue-t-on ça ou un problème de compréhension ?
Pratiques de notation	Plusieurs objectifs évalués = plusieurs notes ?	Le seuil du 4 est posé en référence aux minima requis dans chaque exercice		
		L'échelle est linéaire avant et après le 4		

Cette analyse au travers des catégories conceptualisantes donne à voir sur quelles dimensions de sa pratique Philémon, avec l'aide de ses pairs et du formateur, porte ses réflexions.

Notons d'emblée que les éléments qu'il a présentés dans les deux séminaires n'ont suscité que peu de discussions avec ses pairs. Au contraire. Les principaux aménagements qu'il a réalisés (une page de garde avec une échelle posée en amont, une grille de feedback centrée sur les apprentissages et une pondération explicite référée aux contenus) ont totalement convaincu ses collègues.

D'autres changements significatifs sont à l'œuvre entre la première et la seconde épreuve :

- Une homogénéisation de la complexité des tâches : en effet, il y a une réorientation des choix de Philémon à ce niveau entre sa première épreuve « *qui est en rapport avec ce qu'ils faisaient d'habitude* », sa version revue et la seconde épreuve, dans laquelle ne figurent que des problèmes complexes. Cette « *montée en difficulté* » des tâches qu'il plébiscite et dont il parle souvent au début de la formation semble s'être atténuée ;
- Le retour à l'élève ne se réalise plus seulement au travers d'une note, mais également sous forme d'appréciations « *même quantitatif, avec toujours, parfois, jamais* » et de

¹⁴⁵ Nous observons toutefois que suite au séminaire 2, Philémon est revenu sur cette idée en divisant l'épreuve originale en deux parties, l'une portant sur les apprentissages simples et l'autre sur les plus complexes.

commentaires. Là, il est intéressant de noter que la cohérence de ces éléments de feedback avec les objectifs n'est pas discutée, mais souvent justifiée dans les séminaires.

En revanche, des discussions ont été menées suite au questionnement du formateur. Elles portaient sur trois points, présents dans le tableau, qui restent selon nous ouverts :

- Le niveau des objectifs issus du PER varie et leur exploitation n'est pas toujours simple à comprendre. En effet, la première épreuve se réfère aux composantes des objectifs, la seconde aux éléments de progression des apprentissages. Pourtant, Philémon défend l'idée qu'il enseigne et évalue des compétences, qui font appel à une mobilisation de ressources multiples. À ce titre, il assume complètement une circulation des contenus et des objectifs dans plusieurs exercices, et peut tout à fait en évaluer plusieurs en n'attribuant qu'une seule note au tout. Cette logique, formellement, renverrait à une exploitation des composantes des objectifs, plus globales, mais elle ne semble pas systématiquement appliquée ;
- Les pratiques de pondération, si elles sont désormais pensées en référence à la complexité des contenus dans les tâches, peinent encore parfois à être transparentes (dans la seconde épreuve, elle est par ailleurs fortement discutée, notamment au niveau du total alloué au respect de la propreté). Ce changement permet à Philémon de penser le seuil du 4 en référence aux objets mathématiques, mais la prégnance des points l'engage encore et toujours dans une logique de compensation où, finalement, des acquis chez l'élève peuvent s'annuler avec certaines de ses difficultés. Le problème de Philémon à s'engager dans une évaluation critériée s'appuie sur cette certitude chez lui que les contenus mathématiques circulent partout, et qu'il est dès lors impossible de les hiérarchiser pour déterminer des incontournables en évaluation : *« j'arrive pas à privilégier tel ou tel savoir-faire, tout a de la valeur, mais pas la même »*. Pourtant, lors du séminaire 3, et face à une interpellation d'un de ses collègues, il concède que lorsqu'il cherche les points minimaux dans chaque exercice en référence aux contenus pour déterminer le seuil de suffisance, il est bien dans une perspective de *« critères cachés »*. Toutefois, abandonner les points est pour lui inconcevable ;
- Le rapport entre les objectifs et les tâches demande clarification. On en veut pour preuve la discussion âpre menée lors du séminaire 3 autour de la progression d'apprentissage du PER relevant de l'anticipation des résultats : personne n'est tombé d'accord sur ce que sous-entend cette anticipation d'un point de vue d'habileté cognitive.

Nous observons donc des mouvements importants et rapides dans les pratiques évaluatives de Philémon durant la formation, qui font peu objet de débat au sein des collègues. Nous faisons ici l'hypothèse que c'est notamment son acuité sur les objets mathématiques qui permet ce positionnement. En effet, Philémon, depuis le départ, fait régulièrement appel au sens mathématique des objets et à leurs caractéristiques, est en mesure d'explicitier les contenus de chaque exercice, fait preuve d'un regard critique et assumé sur les contenus du PER.

Mais, ici aussi, ses réflexions semblent se heurter à l'écueil de la notation référée à l'apprentissage. C'est là que le raisonnement engagé traduit une incompatibilité entre les objets mathématiques et leur possible traduction dans des critères à des fins de notation qualitative. La réflexion, dense entre lui et ses collègues, progresse notoirement autour de la pondération au service de la pose du seuil du 4, mais reste ouverte quant à la construction des autres seuils, finalement linéaires. Toutefois, le bénéfice de ce travail se traduit par une page de feedback que Philémon considère comme un acquis dans ses pratiques.

À la lumière de ces éléments, nous sommes en mesure de formuler quelques hypothèses interprétatives traduisant des mouvements dans les pratiques évaluatives de Philémon, et qui nous permettent de voir dans quelle mesure il parvient à travailler leur cohérence du point de vue du modèle théorique :

- Le contenu du PER peut être analysé de manière critique ;
- Les tâches gagnent en homogénéité du point de vue de leur complexité ;
- La pondération, sous forme de points, se réalise en amont et se réfère aux contenus évalués ;
- Le seuil du 4 suit cette logique de recherche de minima dans les exercices en référence aux contenus.

Conjointement, il se dégage également des tensions ou difficultés à dépasser des ruptures d'alignement curriculaire élargi :

- Les objectifs du PER sont posés encore après coup, ce qui tendrait à dire qu'ils ne sont pas une référence pour déterminer le « quoi évaluer ». De manière plus large, nous pouvons nous demander s'il ne faut pas comprendre cet état de fait comme significatif d'une difficulté à exploiter le PER comme un référentiel d'évaluation ;
- La complexité et la circularité des objets mathématiques interfère avec un raisonnement qualitatif sur la pondération et la notation ;

- La pondération se réfère parfois à des contenus dont le rapport avec les apprentissages en jeu est discutable ;
- La notation permet encore des phénomènes de compensation.

Après la formation

4.1.4 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Philémon à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi à la fin de la formation

Rappelons-le, Philémon n'a produit aucun écrit. Dès lors, le tableau ci-dessous rend compte des éléments significatifs d'une seule source de données relativement aux points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre : le second entretien. Elle est directement mise en rapport avec les catégories conceptualisantes, de manière à rester synthétiques et centrés sur nos questions de recherche. Certes, des redondances peuvent apparaître, mais elles doivent à nos yeux être comprises et interprétées comme des signes d'un processus complexe d'appropriation de divers éléments par Philémon. Nous commentons ces données plus bas en les mettant en dialogue avec les autres données.

Tableau 12 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Philémon à la fin de la formation

Données	Entretien 2
Catégories	
Rapport au référentiel	Pas de problème à paraphraser le contenu du PER (106)
	« il faut partir d'une lecture du plan d'études, comme allant chercher « qu'est-ce que je suis vraiment en train de faire ? » (110)
	« y avait une tentation de vouloir panacher, et là, je suis convaincu que... C'est pas pertinent, ce panachage-là, ouais. Ça l'est éventuellement en onzième, quand on veut préparer des élèves à... au panachage de l'examen » (69)
	Une évaluation en cohérence avec le PER et le matériel (174)
	« par rapport à ce référentiel du plan d'études : est-ce que je suis dans un panachage, est-ce que je varie les compétences mathématiques travaillées ? Ou est-ce que je suis, essentiellement, avec des compétences qui sont très similaires aux, à la précédente, mais dans un contexte différent ? » (175)
Expression de l'objet	Evaluer moins de contenu (25) : se demander « quel est l'essentiel de ce thème-là ? » (26)
Elaboration des tâches évaluatives	« L'exercice, c'est le contexte que je donne à l'élève pour vérifier ce que j'ai envie de vérifier. Donc effectivement, le, le choix de la tâche demandée... Ouais, il est indispensable » (64)

	« je pense que tous les, dans tous les thèmes que j'ai l'habitude d'évaluer, de la manière dont j'évalue, y a ces gestes-là, ouais. Et, d'ailleurs, souvent, ben y a cette démarche, qui est de dire « on fait un test assimilé avant le test significatif pour préparer, mais on veut quand même vérifier ça et ça... [...] et pour finir il dure quarante-cinq minutes... » [...] « Et on le revérifie après coup. On leur redonne des points, etc. Donc il est évalué une fois... [...] C'est ça qui me dérange, ouais, y a redondance [...] Et... et en fait, on serait peut-être meilleur avec un petit test formatif de quinze minutes, avant, pour vérifier, et après on teste les deux éléments en cohérence, dans une épreuve » (84-88)
Pratiques de pondération	« valoriser ce qui était le plus ambitieux dans ce qui a été enseigné » (27)
	« dans l'élaboration de l'évaluation, je vais devoir être plus proactif dans l'attribution des points » (31)
	« je suis pas convaincu que... de faire fi des points... ça va me permettre d'être meilleur dans ma cohérence ». « L'acquisition de ces gestes, ben en lui attribuant beaucoup de points, quoi, en lui attribuant des points, et puis en mélangeant tout ça et en ayant ce pot-au-feu, voilà... C'est confortable. Et je pense... voilà, j'arrive pas à me dire que c'est faux » (14)
	La compensation est réaliste, utile à l'élève
	Bénéfice du feedback verbal, mais en même temps, quelle est la plus-value des critères ? (17)
	« pour que l'évaluation, elle soit correcte, il faut que dans la pondération, dans l'attribution de la note, on soit respectueux du degré de complexité » (171)
Pratiques de notation	Les exercices comme des éléments de construction de la note (35)
	« chaque fois je lui produisais des échelles où, effectivement, j'allais chercher en me disant « ben tiens, qu'est-ce qui est utile pour moi ? A partir de quand est-ce que je peux dire que c'est réussi ? etc. » » (39)
	« actuellement, dans la construction de la note, l'attribution de la demi, c'est, c'est... Je trouve que c'est difficile, au bout d'un moment, d'être assez pointu là-dessus, et puis... C'est triste mais, quoi, non, c'est pas triste, mais effectivement, c'est dommage qu'on n'ait pas pu garder ce 4, 5, 6. » (39)
	« C'est un changement, dans le sens oui. Là, on vient de faire une évaluation après, et on a construit cette échelle en amont... et annoncé sur l'épreuve. Par contre, c'est un changement qui a pas coûté » (41-42)
	« C'est très clair de, ce que j'ai envie de valoriser dans les acquis des élèves » (53)
	« en attribuant plus de points à ça [ce qui est important], je sais que quand ça va apparaître en points cumulés, ça va obligatoirement... mériter la suffisance, ou... satisfaisant ou excellent » (53-54)

Ce tableau nous permet de formuler des hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi :

- Le PER semble devenir une référence, surtout au niveau de l'articulation des contenus et des objectifs. Philémon, ici, dit être en mesure de réfléchir à leur éventuelle organisation pour penser son enseignement et ses pratiques évaluatives au travers de reformulations qu'il sait justifier ;
- Les objets mathématiques sont analysés à l'aune de leurs caractéristiques essentielles ;
- Les tâches sont perçues du point de vue de leur complexité, et Philémon est ouvert à les différencier dans les épreuves (TA vs TS) ;
- La pondération peut avoir une dimension qualitative au travers des feedbacks, même si les points restent la référence ;

- Les seuils sont pensés en références aux attentes liées à l'apprentissage, en amont, notamment celui du 4. La notation est en rapport étroit avec la pondération.

S'en dégagent également des tensions, ou difficultés à surmonter des risques de ruptures d'alignement curriculaire élargi :

- La compensation est un phénomène accepté ;
- La notation référée à l'apprentissage est problématique, notamment à cause de la multitude des échelons (onze).

Nous observons donc ici que des mouvements significatifs semblent se confirmer dans les pratiques évaluatives de Philémon. Il le dit explicitement, ils ont pu avoir lieu grâce à l'appropriation du modèle théorique : « *cet alignement curriculaire qui... voilà, d'avoir mis rapidement en exergue la rupture, qui arrive au moment de l'évaluation, de la prise de l'échelle, etc.* » (1). Il a pu notamment mettre des mots sur des phénomènes qu'il rencontrait : « *moi-même dans ma pratique, ben j'ai aussi... de par ma manière d'évaluer, et de par le système d'échelle à points etc., où y avait une rupture* » (1) [...] « *Là, je récupère un peu de cohérence* » (2). Le modèle lui permettrait donc de dire ce qui ne va pas dans une épreuve, elle-même contextualisée dans un processus d'enseignement et d'apprentissage. Certes, son raisonnement sur les ruptures s'ancre ici autour de la notation. Mais il va également au-delà, lorsqu'il explicite une recherche de qualité dans l'évaluation en construction : « *on a gardé cette habitude-là, de rappeler un peu les points qui vont être évalués, qui ont été travaillés, d'avoir l'échelle, mais surtout de garder une place pour dire ce qui a joué et ce qui a pas joué. Et dans cette place de feedback, ben y a effectivement... la recherche des critères... prépondérants de ce qu'on va chercher à vérifier, en termes d'acquisition dans l'évaluation* » (2). Et là aussi, une opérationnalisation du modèle semble en œuvre : « *Voilà, et ça je pense que c'est un apport durable dans la construction de mes évaluations. Cette page de feedback qui passe aussi par la... l'apport d'un tableau, et dans ce tableau les critères... qui sont en lien avec les objectifs, l'enseignement, et la pondération* » (2).

Malgré les tensions soulevées, nous faisons l'hypothèse que l'alignement curriculaire élargi apparaît ici comme un outil analytique compréhensif des contenus à évaluer, permettant non seulement de modifier certaines dimensions de la pratique, mais aussi de mettre des mots sur des phénomènes que Philémon rencontrait et peut maintenant expliciter. Des résistances perdurent encore, mais elles peuvent être en partie contournées au travers de compromis gérables.

Tout au long de la recherche

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse que Philémon est davantage outillé pour éviter des désalignements entre les objectifs, les tâches et, dans une certaine mesure, les systèmes de pondération en points mis en œuvre. Cette hypothèse s'appuie sur l'observation suivante : à la suite du premier séminaire lors duquel Philémon a été confronté au modèle théorique et aux premières réflexions sur la cohérence de l'évaluation sommative avec ses pairs et le formateur, des changements significatifs ont été opérés depuis ses pratiques habituelles.

Nous pouvons dès lors appuyer nos hypothèses interprétatives par trois mouvements importants en œuvre dans les pratiques évaluatives sommatives de Philémon, que nous considérerons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi :

- L'élaboration des évaluations poursuit une recherche de cohérence entre les différentes composantes du modèle théorique de l'alignement curriculaire élargi : *« je pense clairement, dans la construction de l'épreuve, que... ça c'est vraiment un apport de ta relecture... »* (25),;
- Le rôle indispensable du feedback à côté de la note est soulevé : *« Et dans cette place de feedback, ben y a effectivement... la recherche des critères... prépondérants de ce qu'on va chercher à vérifier, en termes d'acquisition dans l'évaluation »* (2) ;
- La construction des seuils a changé, comme nous l'avons abondamment documenté.

D'un point de vue général, Philémon dit que l'alignement curriculaire élargi lui permet d'explicitier certains phénomènes ressentis jusqu'alors. L'analyse de nos données va dans ce sens. Il est toutefois encore en proie à des ruptures de cette cohérence au niveau de la notation, restant convaincu que la pondération en points est un « pis-aller » acceptable. Ce parcours, selon nous, interroge l'opérationnalisation du modèle théorique avec certains objets mathématiques, et donc avec certaines disciplines. Nous y reviendrons dans notre dernière partie.

Mais qu'en-est-il maintenant du point de vue du développement potentiel des compétences évaluatives de Philémon, au regard de ces mouvements de pratique contrastés ?

4.2 Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Philémon

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Philémon au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe.

Il nous semble ici important de reprendre chacune des six compétences et d'en justifier le possible développement. Pour éviter de charger le texte, nous rappelons qu'elles sont décrites dans le portrait de Fernand qui ouvre ce chapitre. Aussi, en nous appuyant sur nos résultats précédents, nous pouvons documenter quelles compétences Philémon¹⁴⁶ semble avoir développées :

Tableau 13 : Indices de développement potentiel des compétences évaluatives de Philémon

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Oui, même si le statut des composantes et des progressions des apprentissages reste parfois à clarifier	Oui, en rapport avec les objectifs dans une perspective analytique			
B	Oui, traduisent des objets clés de la discipline présents dans le PER	Oui, traduisent des objets clés de la discipline	Oui, sont pensées notamment en référence à leur niveau de complexité		
C	Oui, les points tendent à être donnés en référence aux contenus issus du PER	Oui, il y a un souci d'attribuer des points depuis des objectifs portant sur des objets		Plus ou moins, les points perdurent mais permettent une réflexion sur les exigences	
D	Oui, la note a un rapport avec le PER	Oui, la note rend compte d'éléments de	Oui, la note tend à traduire la complexité	Non, la note est issue d'un total de points. Seul	Non, les autres notes sont construites dans

¹⁴⁶ C'est en triangulant nos données et en observant la récurrence de certains propos ou pratiques que nous avons validé, ou non, le développement des compétences au travers de leurs multiples dimensions.

		contenus	des objets présents dans les tâches	le seuil du 4 est estimé en référence aux apprentissages	une perspective linéaire, compensations possibles, ruptures présentes
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures	Ruptures
F	Oui, rapport critique	Oui, choix négociés	Oui, changements assumés	Non, pas d'utilisation d'un système critérié, jugé non meilleur que le système par points	Oui, les pratiques sont justifiées, font l'objet de compromis jugés acceptables

Nous l'avons montré dans notre partie conceptuelle sur l'AL, la perspective que nous en adoptons se traduit au travers de cinq caractéristiques, que nous reprenons pour montrer plus finement le développement des compétences évaluatives de Philémon. Pour la clarté du texte, nous les récrivons en italique dans la section suivante. Il nous semble intéressant de mettre ces différentes caractéristiques en dialogue avec nos résultats pour donner à voir le processus de développement des compétences de Philémon.

4.3 Le processus de développement des compétences évaluatives chez Philémon au regard des caractéristiques de l'assessment literacy

1. *La prise en compte du contexte dans lequel s'inscrivent ses pratiques.* Le contexte prégnant ici chez Philémon est essentiellement relatif à ses élèves. Etant donné qu'il doit les amener vers une transition école-métier, il est particulièrement attentif à ce que son enseignement soit en rapport avec les attentes que ses élèves devront satisfaire. Cette référence pourrait expliquer dans une certaine mesure sa distance critique au PER, mais en même temps, nous faisons l'hypothèse qu'elle pourrait servir sa recherche de justesse en évaluation, pour amener le plus grand nombre d'élèves à leur meilleur niveau.
2. *La négociation par Philémon des objets disciplinaires en jeu dans l'enseignement, l'apprentissage et l'évaluation.* Philémon choisit les objets disciplinaires qu'il va traiter, et surtout la manière avec laquelle il va les enseigner, en fonction de plusieurs facteurs relatifs à cette transition vers le monde professionnel et pensés de plus en plus en référence à l'alignement curriculaire élargi. Ce processus décisionnel, qui caractérise l'AL, nous fait penser à une action de négociation que Philémon mène à chaque fois qu'il s'agit de savoir ce que les élèves vont devoir apprendre.

3. *Sa prise en compte des processus d'enseignement et d'apprentissage pour penser les pratiques évaluatives.* Dès le départ, Philémon accorde une importance au rapport entre ses pratiques évaluatives et les processus d'enseignement et d'apprentissage. Ce qui évolue, c'est ce sur quoi porte sa réflexion. Alors qu'au début il est davantage conditionné par les moyens et les tests « métiers », c'est en référence aux objectifs et aux contenus qu'il se réfère progressivement pour organiser les activités en classe puis dans l'évaluation sommative.
4. *Les connaissances incontournables en évaluation de Philémon référées à son contexte.* Philémon parle de la taxonomie comme une connaissance lui permettant d'être plus explicite sur le niveau de complexité de ses épreuves : « *c'est ce qui me donne envie de mieux distinguer TA, TS, qui est effectivement de... de garder l'évaluation significative pour... évaluer le degré de complexité plus important* » (entretien 2, PH 130). Nous faisons également l'hypothèse que le modèle de l'alignement curriculaire élargi lui a permis d'affiner sa connaissance du PER, des caractéristiques des tâches évaluatives, et des systèmes de pondération, au vu de ses propos mais surtout des épreuves créées. En revanche, ses connaissances en notation - comme celles relatives à la nature des échelles - semblent encore en construction, notamment parce qu'elles se heurtent à des difficultés de considérer les critères comme des outils pertinents en mathématiques.
5. *Une pratique évaluative considérée comme relevant, pour Philémon, d'un compromis permanent entre différentes dimensions.* Ces compromis sont très explicites lorsque Philémon exprime ses impossibilités à mettre en œuvre une évaluation critériée et une notation uniquement basée sur des critères. Nous observons concrètement ces compromis dans les deux systèmes de pondération et de notation qu'il élabore dans ses deux épreuves, lorsqu'il est prêt à construire un seuil de suffisance en fonction des apprentissages, mais pas pour les autres notes. D'autres compromis sont également présents dans son rapport au PER, lorsqu'il reformule des objectifs, ou quand il décide de l'importance des objectifs à évaluer.

4.4 En guise de synthèse provisoire

Au regard de ce qui précède, nous voyons que c'est la compétence A, relative à l'appropriation du contenu du plan d'études et des objets de savoir, qui semble se développer principalement, malgré les nuances que nous avons relevées.

La compétence B, centrée sur la conception des tâches évaluatives, est également

significativement en développement.

En revanche, c'est au niveau des compétences C, mais surtout des D, E et F que des indices empêchant leur développement apparaissent. En effet, elles font toutes les quatre références à des dimensions relevant de la pondération et de la notation, celles-là même qui sont sources de tensions dans les pratiques de Philémon. À l'instar de la situation de Paul, nous observons donc ici une « zone de perturbation » dans le développement des compétences évaluatives de l'intéressé, mais pour d'autres raisons que son collègue. Malgré ces différences, il devient délicat de documenter, au regard de nos données, un développement significatif à ce niveau.

Ici encore, en référence à notre perspective théorique d'AL, nous observons qu'une tension présente sur une des catégories entraîne de fait une résistance au développement de plusieurs compétences. Il en va ainsi de la question de la pondération et de la notation, dont nous avons déjà parlé. Ce constat revient encore une fois à nous demander s'il est justifiable de considérer qu'une évaluation curriculairement alignée entre les objectifs et les tâches, mais en rupture partielle avec la pondération et plus conséquemment au niveau de la notation, peut être considérée comme cohérente ? Comme nous l'avons dit avec Paul, nous y reviendrons dans notre conclusion.

Philémon semble donc avoir acquis des connaissances et des compétences clés en s'appropriant le modèle de l'alignement curriculaire. Dans son acception élargie ? Il est encore ici une fois difficile de l'affirmer. Nous pensons toutefois qu'une posture réflexive est en voie de développement à certains niveaux grâce au processus de formation, au cœur de laquelle son jugement professionnel a gagné en cohérence grâce à la référence permanente au modèle. Sa non-tentative de système de pondération critérié en est selon nous un exemple pertinent, ce renoncement n'étant pas synonyme d'immobilisme. Sa situation est toutefois interpellante, d'un point de vue de formation : s'il prend conscience rapidement d'un certain nombre de phénomènes et semble s'approprier facilement certaines composantes de l'alignement curriculaire élargi pour penser ces derniers et modifier ses pratiques, d'autres semblent ne pas l'aider. Des arrangements prennent alors le dessus, mais Philémon peut maintenant les assumer. Il semble avoir comblé ses attentes de départ pour évaluer de manière plus « juste », au-delà des dilemmes persistants.

5 Portraits 4 et 5 : Anouck et Alizée, enseignantes de français

La réalisation des portraits d'Anouck et Alizée se conjuguent, car ils s'articulent autour des mêmes épreuves. Nous tentons donc ici d'éviter les répétitions, tout en mettant en évidence la singularité du cheminement respectif de chaque enseignante. Ces portraits s'appuient sur les données suivantes :

- Leur premier entretien et leur premier écrit, réalisés tout deux avant la formation. Cette partie donne à voir respectivement d'où elles partent, quelles sont leurs interrogations ;
- Leur première et seconde évaluation¹⁴⁷ ainsi que leur discussion proposée par le formateur, leur deuxième écrit, tous produits durant la formation. Cette partie est commune aux deux enseignantes, comme la suivante ;
- Les échanges qu'Anouck et Alizée ont eus avec leurs pairs et le formateur durant les séminaires de formation 2 et 3 ;
- Leur second entretien et leur troisième écrit, réalisés à la fin de la formation. Cette partie dissocie également les propos des deux enseignantes.

Abordons maintenant les dimensions interrogées par nos questions et sous-questions de recherche.

5.1 Les dimensions des pratiques évaluatives d'Anouck et Alizée sur lesquelles elles portent chacune leur attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi.

Avant la formation

5.1.1 Les dimensions de la pratique évaluative sur lesquelles chacune porte son attention *avant la formation*

Anouck

Un important questionnement relatif aux pratiques évaluatives sommatives est exprimé par Anouck dans son premier entretien et son premier écrit intitulé « L'évaluation ? Entre illusions, désillusions et magie... ». Ce n'est pas anodin. En effet, elle y exprime plusieurs doutes. Tout d'abord, qu'évaluer ? Par exemple, relativement à la production textuelle, elle dit : « *au moment*

¹⁴⁷ Anouck a réalisé ses deux épreuves en collaboration avec sa collègue directe, Alizée, qui participe également à la recherche et dont le portrait suit celui-ci.

où il s'agit de créer un test, de le faire passer puis de le corriger, les doutes surviennent. [...] que teste-t-on ? Le portrait ou le génie de l'écriture ? [...] Que devais-je évaluer ? Leur génie littéraire ou ce qu'ils ont appris ? ». Cette interrogation sur l'objet évalué revient dans l'entretien : « C'est aussi, des fois, qu'est-ce qu'on attend... enfin, je me rends compte qu'on sait pas toujours ce qu'on attend, finement, quand on fait un test, enfin, qu'est-ce qu'on veut ? Si on veut juste qu'ils sachent écrire, ou bien, s'ils doivent avoir atteint des objectifs plus précis à l'intérieur ». En d'autres termes, ce qui pose problème à Anouck est la justesse de ses pratiques, le fait d'être au plus proche d'une recherche de ses attentes. Elle écrit : « En résumé, ce qui me perturbe avec l'évaluation, c'est cette difficulté à me dire : j'ai fait juste, j'ai bien fait mon travail », propos qu'elle étaye dans son entretien, très tôt : « je trouve que c'est aussi ça, c'est cette légitimité de l'évaluation quoi ; comment on peut dire que notre évaluation elle est juste, elle est correcte, pis que la note, donc la moyenne qu'ils vont avoir, elle serait la même avec un autre prof, ou dans d'autres conditions, et pis que vraiment elle reflète leurs compétences dans une branche ». Elle questionne ici la validité de son évaluation. Toutefois, pour elle, ses doutes se relativisent avec l'expérience : « Il y a rarement une incompréhension complète à propos d'un résultat [...] la magie opère ». Elle va même jusqu'à affirmer que finalement, « on retombe quand même souvent sur nos pieds ».

Mais le questionnement est bien là, et Anouck exprime un besoin d'outils pour valider ses manières de faire : « j'ai besoin, en fait, d'avoir des outils pour m'assurer que ma manière de fonctionner est, en tout cas, cohérente », pour « s'assurer qu'on est dans la bonne direction, en fait ».

Ensuite, lorsqu'elle décrit ses pratiques évaluatives, plusieurs éléments semblent relativement stabilisés. Ils sont exprimés lors de l'entretien :

- Les objectifs sont donnés au préalable, mais ils sont reformulés de manière très précise : « parfois, je me demande si c'est trop ». Toutefois, ils semblent avoir un statut particulier, notamment ceux du PER mentionnés sur les évaluations : « je fais toujours les objectifs après ; je fais d'abord le test, pis après je fais les objectifs, mais dans ma tête, je me suis déjà... en français, je me suis demandé ce que je voulais tester ». Les objectifs sont très précis lorsqu'elle évalue la technique de la langue, plus globaux lorsqu'elle est sur des productions ou de la compréhension ;

- Il en va de même pour les critères et les grilles d'évaluation, également transmis à l'avance aux élèves. À ce niveau Anouck dit : *« j'ai assez peu de problèmes pour poser mes critères, parce que j'ai une idée du texte final que je veux »* ;
- Elle paraît passer assez systématiquement par une analyse de l'objet de savoir pour préparer ses évaluations. Quand elle exemplifie avec la production textuelle, elle explique : *« C'est un peu comme si je prenais un texte, je me dis « qu'est-ce qu'il y a dedans ? », pis je le démonte... »* ;
- Les évaluations, bien que pensées en début de séquence d'enseignement, sont adaptées au fur et à mesure que les apprentissages ont lieu avec les élèves : *« on peut justement faire le point de ce qu'on a vu, de s'assurer que vraiment les critères qui sont dans notre grille, on les a vus au préalable en classe »* ;
- Un TS est toujours un mélange de technique et de production.

Nous observons donc des éléments contrastés dans les pratiques exprimées en début de formation. Qu'en est-il plus précisément ?

Relativement au PER, Anouck semble critique, voire parfois désemparée. Elle l'exploite comme fil rouge pour sa planification annuelle, mais au niveau de l'évaluation, cela semble plus délicat : *« je l'utilise très très peu pour construire l'évaluation quoi, je m'en rends bien compte »*. Elle va même plus loin, en affirmant : *« mais j'ai pas l'impression que ces critères [du PER] m'aident beaucoup [...] Je sais pas comment faire le lien entre un objectif général du PER et puis une évaluation d'un texte écrit quoi, je vois pas »*. Cette réflexion compte surtout pour les objectifs complexes hors fonctionnement de la langue qu'elle pourrait *« tordre dans tous les sens, ça irait quand même »*. Pour conclure le sujet, elle dit d'ailleurs : *« J'ai l'impression que c'est toujours les mêmes, de toute façon, donc voilà... »* ;

En ce qui concerne les tâches évaluatives, d'autres éléments semblent s'inscrire pleinement dans ses pratiques. Il est intéressant de constater qu'à ce niveau, Anouck fait des différences entre les objets qu'elle évalue : *« pour une compréhension de texte, par exemple, où y a pas de critère, c'est des questions, je vais reprendre, souvent, une même structure que ce qu'on a vu avant ; donc des questions de compréhension, des questions générales de compréhension, des questions de vocabulaire... »*. Pour des évaluations de production écrite, des critères d'évaluation sont exploités, des éléments de fonctionnement de la langue sont introduits *« j'ai une partie « technique de la langue » qui est toujours la même, en fait. Donc ce qui va varier... enfin, elle peut un peu évoluer en fonction de ce qu'on voit, mais après c'est orthographe,*

conjugaison... enfin, des choses comme ça ». Mais elle pense toujours aux attendus en matière de production des élèves, relativement à l'objet de savoir évalué : « *je réfléchis en matière* ».

Ses pratiques de pondération, quant à elles, semblent également stables : « *j'ai jamais un total de points auquel je veux arriver, mais je vais regarder chaque question, un peu comme je formule les questions, j'attends un certain nombre d'éléments dans la réponse, et je vais mettre les points par rapport à ça* ». On le voit, une réflexion entre l'attribution des points et le contenu, via les tâches évaluatives, semble déjà à l'œuvre, hors des contraintes arithmétiques : « *Ça reste quand même toujours à peu près au niveau du même nombre de points, par je ne sais quel hasard, mais, je veux dire, mais c'est vrai qu'effectivement, je fais jamais attention à ça* ». Là, elle dit pondérer davantage les apprentissages plus complexes, et afficher clairement l'attribution des points sur les épreuves, à côté des critères lorsque c'est le cas.

Mais si un raisonnement existe déjà entre la pondération et la notation au niveau de la création du seuil du 4 « *je fixe mon 4 en pourcentage, mais je compte les points minimaux qu'ils devraient avoir pour moi, avoir acquis la matière* », de gros doutes et des pratiques en questionnement prennent le dessus : « *le reste, c'est du brochage, et encore une fois, suivant ce que je mets un petit poil en dessus, un petit poil en dessous, ben...* ». Pour Anouck, à ce niveau, une certitude semble ancrée : « *on peut faire un test, et pis finalement, c'est l'échelle qui va poser la note plus que le test* ». C'est peut-être une des raisons pour lesquelles nous retrouvons trois axes de réflexion à ce niveau chez Anouck, traduisant, selon nous :

- Une ambivalence autour du moment où l'échelle est créée : « *je fais pas toujours l'échelle vraiment complètement avant, par contre, je fais jamais le total de points... [...] Donc parfois, je fais l'échelle complètement avant, c'est rare mais ça arrive, surtout pour les petits tests ou des choses comme ça, et puis, parfois je fais le test, je laisse répondre les élèves, je fais l'échelle quand même avant de corriger, en fait. Donc ça dépend, mais... [...] ça me convient pas de faire l'échelle après [...] J'ai pas une règle fixe, pas du tout* » ;
- Une difficulté à justifier l'exploitation de barèmes : « *[le 4] c'est toujours aux alentours... entre 60 et 75%, enfin, 70, je dirais. En fonction de la difficulté du test* », tout en expliquant les raisons qui lui font renoncer au barème fédéral : « *si je voulais prendre ce barème-là, je devrais concevoir mon test différemment. Tandis que là, j'ai l'impression que j'ai au moins la liberté ensuite de faire correspondre l'échelle à la difficulté de mon test* » ;

- Un attachement à la régularité des échelles qui manque d'arguments pédagogiques : « dans la loi, on nous a dit qu'on devait avoir des échelles régulières », « *je travaille avec la régularité, donc je compte les points pour que ce soit régulier.* ». Pour elle, la régularité renvoie donc à une forme de justesse : « *moi c'est comme ça que je le concevais [régulier sous le 4, plus large entre 4 et 6], et c'est pour ça qu'avant, j'ai jamais pu faire une échelle qui soit d'un point en dessous à chaque fois, pis que tout d'un coup elle soit très large* ».

Globalement donc, Anouck est en grand questionnement. Attentive à une recherche de cohérence entre enseignement et évaluation, convaincue que l'enseignement du français est « *spiralaire* » et qu'à ce titre, il est délicat pour elle de déterminer quelles compétences évaluer dans des situations complexes, elle est attentive aux objets de savoir que ses élèves doivent maîtriser. À ce titre, elle tente de mettre en œuvre des pratiques les plus cohérentes possible, mais est également prise dans des doutes importants, notamment au niveau de la notation et de l'exploitation du PER.

Alizée

Alizée écrit son premier récit de pratique trois semaines avant le premier entretien. Ce texte est déjà porteur d'interrogations qui portent sur la collaboration et l'impact, positif ou négatif, que cette dernière peut avoir sur la qualité des épreuves. En effet, nous observons qu'elle y questionne la perception des attendus de chacun et leur possible conjugaison au niveau des contenus à évaluer, du choix des tâches et des échelles. Mais on y observe déjà une attention particulière à élaborer des épreuves les plus justes possibles, attention qui se retrouve dans son entretien.

Là, son questionnement est d'entrée ciblé. Elle se demande par exemple comment évaluer certains contenus : « *je me trouve devant un peu une question : comment faire pour évaluer ça [l'expression orale] ?* ». Elle s'interroge également sur ses grilles critériées, qu'elle utilise déjà en production écrite : « *continuer à réfléchir là-dessus, quel poids on met à chaque partie, le fond, la forme, à quel point on introduit aussi de la structuration dedans, enfin, ça nous on le fait, mais... voilà, quel équilibre on donne à chaque partie* ». Nous voyons ici que le questionnement touche les contenus et la pondération, mais également la construction de la note, qui est, pour elle, source de tensions : « *comment et quand on fixe nos échelles ? On fait pas toujours pareil : parfois, on fixe l'échelle avant, parfois après, parfois on se dit « on choisit*

un pourcentage, mais pourquoi est-ce qu'on le choisit ? ». Enfin, moi j'ai jamais tellement pris le temps de réfléchir à ça ». À ce titre, et lorsqu'elle est amenée à expliciter plus en détail ses pratiques et conceptions, elle dit qu'elle affectionne les échelles régulières sous le 4, mais ne sait pas le justifier. Par ailleurs, elle est ouverte à augmenter le seuil du 4 en fonction de la complexité de l'évaluation, même si elle se demande si c'est juste : « Je trouve que, par exemple, dans un test de grammaire où ils ont dû apprendre, ils doivent appliquer, c'est légitime qu'on soit plus strictes, que le 4 soit plus vers 75% de capacités, enfin, de connaissances, de réussite ; ça je trouve bien. Quand ils doivent faire une rédaction, qu'ils doivent appliquer mes douze critères là – ou je sais pas combien j'en ai –, c'est pas possible, je trouve que s'ils ont réussi déjà à... enfin, là je pense on applique pas 75% quoi, on élargit plus quand y a... on a mis ensemble plus de critères, plus d'objectifs et voilà. [...] Mais, voilà, je sais pas si c'est juste ou pas ». Et cela, même si elle est confrontée au barème fédéral, vanté par la majorité de ses collègues, et face auquel est restée insécure « Pour moi, ça [le barème fédéral] n'a pas tellement de sens. Enfin, je sais pas... ». Des doutes persistent encore sur la manière et le moment de poser le barème : « idéalement, par rapport à un test, ce serait mieux de le faire avant, parce que sinon, on met les élèves là où ils sont, pis on se dit « c'est où qu'ils méritent 4 ? ». Enfin, ça ça veut dire que c'est pas, ça veut dire qu'on évalue pas leurs compétences, mais qu'on se dit « bon là, ça va ». Mais, quand on met l'échelle avant, à part ça, on pousse¹⁴⁸ aussi, parce que, si je mets une échelle comme ça avant pour ma rédaction, et pis que j'ai un élève je me dis « ah, ça ferait combien ? Ah non, allez, je lui trouve deux points quelque part ». Bref, on le voit, pour Alizée, « en tout cas une des grosses questions : cette histoire de « comment fixer mes échelles », enfin, ça ça reste une question (...) ».

Du point de vue de ses pratiques évaluatives, certaines habitudes semblent en revanche stabilisées :

- Une attention particulière est donnée à enseigner, puis à évaluer, des éléments de fonctionnement de la langue au travers des quatre compétences de français (compréhension et production de l'écrit, compréhension et production de l'oral) ;
- Les contenus des TA et TS sont clairs : fonctionnement de la langue pour le premier, tâches complexes et fonctionnement de la langue pour le second : « En général, on essaie d'accrocher nos sujets de structuration avec la séquence qu'on est en train de faire, dans la mesure du possible ». Leur fréquence et leur nombre est également stable ;

¹⁴⁸ Pougnier est synonyme de tricher, en langue vaudoise.

- Les objectifs, reformulés, sont donnés à l'avance aux élèves, comme dans l'exemple écrit dans son texte : « *M. a également tapé les objectifs qui reprenaient ceux de notre séquence pour pouvoir les donner aux élèves une semaine avant le test (procédure que nous employons habituellement pour tous les tests)* » ;
- Toutes les épreuves sont réalisées en collaboration étroite avec Anouck et une autre collègue. À tour de rôle, elles se répartissent le travail de conception, puis confrontent leur analyse afin d'arriver à une épreuve finale.

Alizée est donc ici assez générale dans l'explicitation de ses pratiques, hormis son questionnement sur la notation. Mais plus loin, elle délivre quelques autres éléments de réflexion intéressants.

Relativement au PER tout d'abord, elle dit le traduire pour le rendre compréhensible aux élèves : « *je prends les objectifs du PER, pis je les traduis pour mes élèves : qu'est-ce que ça veut dire, enfin, qu'est-ce qu'il faut qu'ils fassent, en fait. [...]. Enfin là encore, c'est très vague, très global quoi. Donc après je redétaille un peu ce qu'ils doivent faire. [un PER] vulgarisé, je dirais* ». S'il est peu compréhensible, il est par ailleurs peu novateur : « *on travaillait déjà beaucoup avec le PEV¹⁴⁹, nous, en tout cas, donc finalement, ça changeait pas des masses notre travail quoi* ». Le PER ne semble donc pas être une référence pour elle, juste un moyen de contrôler qu'elle ne s'éloigne pas trop des attendus : « *Moi j'aime bien avoir quelque chose comme ça, dans le sens je me dis « bon, voilà, ça on a fait, c'est bien, on est dans la cible »* ».

Au niveau du contenu des évaluations, Alizée dit avoir de grandes difficultés à évaluer le fonctionnement de la langue en production. Notamment pour ne pas pénaliser certains élèves en difficulté à ce niveau : « *en orthographe, on s'est toujours dit « mais on a tous, dans nos classes, des élèves qui ont de la difficulté, on veut pas non plus mettre, dans une rédaction, si y en a qui sont nuls en orthographe, ça doit pas les empêcher de faire une note suffisante* ». Ça c'est évident ». Par ailleurs, elle relève ses difficultés à évaluer parfois certains contenus, qu'elle décrit, sans les exemplifier, comme « *des choses pointues* ». En revanche, et malgré quelques imperfections, ses grilles de critères lui conviennent bien : « *Alors, ça arrive aussi que, des fois, ils aient tout bien rempli la grille, pis que leurs textes soient pas très intéressants. Mais ça me dérange pas tellement, parce que [...] ils ont quand même le droit de faire une bonne note,*

¹⁴⁹ Plan d'Etudes Vaudois, mis en œuvre en 2000, dans une approche par compétences, puis remplacé progressivement par le PER dès 2011.

parce qu'ils ont compris plein de choses ». Leur élaboration se fait en référence à des éléments de contenus qui sont déjà identifiés en début de séquence. À ce titre, les tâches évaluatives sont proches de ce qui a été fait en classe, et elle dit que *« quand c'est des exercices, j'essaie d'en faire un par objectif que j'ai donné... »*. Une attention semble donc exister à mettre en lien tâches et objectifs, et de garder un niveau de complexité élevé *« C'est quand même toujours des choses complexes »*.

Du point de vue de la pondération, ses grilles peuvent être adaptées aux différentes classes et donnent globalement satisfaction : *« j'aime bien les critères, parce que c'est pour le retour pour les élèves, de leur expliquer ce qu'on a jugé »*. Toutefois, Alizée n'est pas au clair sur le niveau de détail de ses critères, sur leur objectivité. Paradoxalement, elle dit que ses critères sont référencés au plan d'études : *« je choisis dans le PER quels sont les critères que j'ai envie d'évaluer dans la rédaction »* alors qu'elle ne le mentionne pas comme référence. Ils suivent également son enseignement, s'inspirent toujours de grilles pré-existantes, et sont souvent très nombreux : *« Ouais, y en a beaucoup »*. Ceux qui portent sur le fonctionnement de la langue sont omniprésents : *« Alors y en a qui sont toujours : orthographe, conjugaison, syntaxe, grammaire, vocabulaire... »*, mais certains changements sont à l'œuvre, en fonction du contenu, comme sur cet exemple de compréhension écrite : *« ils changent aussi suivant le livre qu'on a lu, parce qu'on va pas toujours leur demander la même chose »*. Ces critères sont toujours appariés à des points, pensés en référence au contenu : *« Si je demande une chose précise, je mettrai un point, enfin, je veux dire, si je dis « je veux un narrateur en je », ben j'ai mis un point pour « oui » ou zéro pour s'il s'est trompé. Ça c'est des choses comme ça. Par contre, si je demande quatre événements bien développés, ben je me dis « bon, j'aimerais que ça ait un certain poids » »*. Un équilibre est recherché ensuite entre le poids donné à la structuration et à la production.

Globalement donc, nous observons qu'Alizée fonctionne avec passablement d'habitudes, mais que de multiples questions et tensions sont à l'œuvre. Si certaines dimensions, comme le statut des objets de savoir, semblent peu présentes dans ses réflexions, d'autres en revanche, comme la notation, sont porteuses de fortes attentes en termes de formation.

Nous observons donc que les deux membres de ce duo investiguent différemment certaines dimensions de leurs pratiques évaluatives, mais se rejoignent sur d'autres, comme la notation.

Pendant la formation (Anouck et Alizée)

5.1.2 Les points de cohérence et de ruptures d'alignement observés par le duo dans ses pratiques d'évaluation sommative lorsqu'il se réfère au modèle de l'alignement curriculaire élargi *pendant la formation*

Rappelons ici brièvement les trois temps de la réponse à cette sous-question :

- Présentation des données brutes en nous centrant sur les parties qui ont suscité un débat ou qui mettent en évidence des zones de tensions, de discussions, et en respectant l'ordre chronologique dans lequel elles ont été produites (ici : épreuve 1, épreuve 2, écrit 2).
- Présentation d'un tableau de synthèse de ces éléments que nous interpréterons en référence aux catégories conceptualisantes que nous avons définies.
- Commentaires de nos résultats.

5.1.2.1 Première évaluation créée pendant la formation¹⁵⁰

Cette évaluation porte sur la compréhension de texte (annexe Partie 5, 9.2.1). Elle a été conçue pour des élèves de 9 VP (Voie prégyrnasiale) et vise à être mise en œuvre après le séminaire. Elle est constituée de deux pages de garde. La première présente les trois objectifs évalués dans l'épreuve (observer la position du narrateur et le genre d'un extrait, utiliser le vocabulaire pour comprendre un texte, comprendre un extrait de texte inconnu) reformulés par Anouck et sa collègue à partir du PER¹⁵¹. Ces objectifs sont ensuite déclinés en sous-objectifs, et mis en correspondance avec les références exploitées durant l'enseignement (moyens, ouvrages de base comme le livre unique ou le Bescherelle). Etant donné que des éléments de discussion sont ici déjà proposés par le formateur, nous privilégions une présentation des parties de l'épreuve au regard de ces éléments.

¹⁵⁰ Logiquement, les éléments de présentation et de discussion des épreuves sont les mêmes pour Alizée et Anouck. Ces épreuves ont par ailleurs été mises en œuvre dans leur classe respective.

¹⁵¹ Ce qui confirme les dires d'Anouck exprimés lors du premier entretien.

Objectif 1 : Observer la position du narrateur et le genre d'un extrait :

- repérer la position (= point-de-vue) du narrateur (LU p. 275),
- identifier le genre du texte (Bescherelle collège n° 364-365 et 371 à 374)

Objectif 2 : Utiliser le vocabulaire pour comprendre un texte :

- Identifier et utiliser les reprises nominales et pronominales (LU p.263), et donner leur classe grammaticale (p.297 et suivantes)
- Donner le sens de certains mots, leur synonyme et antonyme (LU p.362) à l'aide du dictionnaire.

Objectif 3 : Comprendre un extrait de texte inconnu :

- repérer le sujet ou le thème d'un extrait donné en observant les champs lexicaux les plus importants (LU p.365)
- en dégager les événements principaux en repérant le schéma narratif d'un extrait donné (Bescherelle collège n°358)
- identifier les personnages et leurs sentiments,
- percevoir l'importance du décor pour les événements,
- percevoir les sensations et émotions que peut ressentir le lecteur.

Figure 42 : Anouck et Alizée, épreuve 1, compréhension de texte, page de garde 1 (objectifs reformulés)

Ici donc, le formateur questionne le processus de reformulation des objectifs du PER. Par ailleurs, il questionne le nombre de points attribués à chaque objectif (3,5 pour le 1, 9 pour le 2 et 17,5 pour le 3). Sous une forme quelque peu provocatrice mais bienveillante, il signale que la cohérence de l'alignement est selon lui davantage respectée entre les tâches et les objectifs reformulés qu'avec ceux du PER. Là, il investit encore les logiques ayant participé à cette reformulation.

Puis, vient une seconde page de garde sur laquelle figurent deux objectifs du PER (L1 31-Lire et analyser des textes de genres différents et en dégager les multiples sens..., et L1 36-Analyser le fonctionnement de la langue et élaborer des critères d'appréciation pour comprendre et produire des textes...), leurs composantes sélectionnées (cinq pour le 31, aucune pour le 36) et leurs progressions d'apprentissages (huit pour le 31, une pour le 36). L'échelle, posée en amont, figure en fin de page.

Date :	Nom, prénom :	9VP/
/30	Français – TS – Compréhension écrite Autour d'Ulysse...	

Objectif L1 31 — Lire et analyser des textes de genres différents et en dégager les multiples sens...

Composantes

- ... en mettant en évidence l'organisation du texte et la progression du récit ou des idées
- ... en utilisant les moyens de référence
- ... en mobilisant et en développant ses connaissances langagières (lexicales, grammaticales,...) et extralangagières (connaissance du monde, références culturelles,...)
- ... en dégagant le sens des mots et des phrases, à partir du texte et du contexte
- ... en dégagant le point de vue de l'auteur

Progressions des apprentissages

- Identification :
 - des sujets ou des thèmes par le repérage des champs lexicaux dominants et/ou spécifiques
 - du genre du texte
 - de l'explicite et de l'implicite
- Interprétation des reprises anaphoriques (pronoms, synonymes, périphrases, termes génériques,...)
- Reconnaissance des idées principales
- Distinction entre l'auteur et le narrateur
- Identification :
 - de la position du narrateur (interne, externe)
 - des éléments du décor
 - des personnages, de leurs relations et de leurs rôles
- Repérage d'une structure narrative (schéma narratif)
- Analyse des sensations, des émotions et des sentiments des personnages
- Distinction entre les valeurs de l'imparfait et du passé simple ou du passé composé

Objectif L1 36 — Analyser le fonctionnement de la langue et élaborer des critères d'appréciation pour comprendre et produire des textes...

Composante
aucune ne convient pour la progression voulue.

Progressions des apprentissages
Compréhension et utilisation des termes métalangagiers (champ lexical, synonyme, antonyme, registre de langue,...)

Echelle :

30-29 : 6	28 ⁵ -27 ⁵ : 5.5	27-25 : 5	24 ⁵ -22 ⁵ : 4.5	22-20 : 4
19 ⁵ -16 ⁵ : 3.5	16-12 : 3	11 ⁵ -6 ⁵ : 2.5	6-0 : 2	

Figure 43 : Anouck et Alizée, épreuve 1, compréhension de texte, page de garde 2 (objectifs du PER et échelle)

Là, les éléments de discussion proposés sont les suivants :

- Une interrogation sur les objets de savoir évalués : « Les caractéristiques principales du genre de texte abordé sont-elles prises en compte ? Que faites-vous, par exemple, de la question de la chronologie ? » ;
- Une question sur la quantité d'éléments annoncés comme évalués : « Il y a beaucoup d'éléments. Sont-ils tous évalués ? » ;
- Un élément de débat relatif à la pertinence d'évaluer le L1-36, ou de le séparer du L1-31 : « Est-ce pertinent ici, vu le nombre d'items qui l'évaluent ? Ou alors vous mettez deux notes ? » ;
- Un questionnement sur la nature de l'échelle : « Pourquoi les échelons s'agrandissent-ils au fur et à mesure que l'on « descend » ? Comment gérer les différents objectifs évalués avec une somme de points ? ».

Le contenu de l'épreuve consiste en une traduction de l'Odyssée de Homère, inédit pour les élèves et long de 69 lignes, qui sert de support à 20 questions mises en correspondance avec les trois objectifs reformulés (quatre questions pour l'objectif 1, cinq pour le 2 et onze pour le 3). Les points et demi-points figurent à côté de chaque question ou sous-question.

Etant donné que les deux éléments de discussion relevés par le formateur concernent toute l'épreuve, nous la reproduisons en intégralité. Ces éléments sont les suivants, et concernent donc les trois parties décrites plus haut :

- Évalue-t-on de la compréhension ou de l'analyse ?
- Comment sont attribués les points ?

<p>Objectif 1 : Observer la position du narrateur et le genre de l'extrait. /3⁵ pts</p> <p>1. Quel est le point-de-vue du narrateur ? _____ Entoure dans le texte un élément qui prouve ta réponse. (_____/1pt)</p> <p>2. Trouve un titre pour cet extrait: (_____/1pt)</p> <p>3. Dans les sous-genres suivants, souligne celui qui correspond à l'extrait donné : (_____/0⁵pt) Récit poétique – récit d'aventures – récit policier – récit de vie – récit noir</p> <p>4. Lis la définition, dans le dictionnaire, du mot « épique », puis explique en quoi cet extrait est un récit épique. (_____/1pt)</p> <p>_____</p> <p>_____</p> <p>_____</p>
--

Figure 44 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 1

Objectif 2 : Utiliser le vocabulaire pour comprendre un texte

/9 pts

1. Dans le texte, souligne :
 - a. La reprise qui désigne Poséidon par une périphrase. (____/1pt)
 - b. La reprise qui désigne Ino par une périphrase. (____/1pt)
2. Choisis la phrase exprimant le sens que les mots suivants ont dans le texte: (____/4pts)
 - a. L. 12, se dérober :
 - ☐ : Il doit travailler avec des copains, mais il se dérobe sans arrêt.
 - ☐ : Il sentit la terre se dérober sous ses pieds.
 - b. L. 19, déferla :
 - ☐ : Les supporters déferlèrent sur le terrain.
 - ☐ : La cavalière déferla sur son cheval.
 - c. L. 54, écume :
 - ☐ : Après l'effort, les naseaux du cheval étaient couverts d'écume.
 - ☐ : Il attend que l'écume se forme pour plonger les pâtes dans l'eau.
 - d. L.61, raboteuse :
 - ☐ : Son grand-père était un vieillard tout raboteux.
 - ☐ : Il s'écorce les genoux sur le plancher raboteux.
3. Indique un synonyme de même classe grammaticale pour le mot suivant : (____/1pt)
 - a. L. 31, déchaîne-t-il : _____
4. Indique un antonyme de même classe grammaticale pour le mot suivant : (____/1pt)
 - a. L. 25, s'y cramponna : _____
5. A quelle occasion pourrais-tu, comme Ulysse, te « rompre les os », l.62 ? (____/1pt)

Figure 45 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 2

Objectif 3 : Comprendre un extrait de texte inconnu**/17⁵ pts**

1. L.18-31: Relève un champ lexical : _____
Souligne au minimum 5 mots ou expressions qui s'y rapportent. (_____/3pts)
2. Complète le schéma narratif suivant : (_____/5pts)

_____	Ulysse a quitté Calypso et vogue sur son radeau.
Complication - perturbation	Poséidon...
_____	Ulysse ...
_____	Au bout de deux jours
Résolution	
Situation finale	Ulysse est sauvé.

3. Quel dieu s'acharne contre Ulysse ? _____
 4. Quel est son attribut ? _____
 5. Quelles sont les déesses qui lui viennent en aide ? _____
 6. De quelle façon chacune d'elles lui viennent-elles en aide ? _____

 7. Qui se métamorphose et en quoi ? _____

 8. Comment Ulysse réagit-il face aux épreuves qu'il doit traverser ? _____

 9. A quoi est-il comparé à la fin du texte et pourquoi ? _____

 10. Quelle est l'importance du décor dans cet épisode ? _____

 11. Quelles sensations et émotions le narrateur cherche-t-il à transmettre au lecteur dans ce passage ? _____

- (_____/9⁵pts)

Figure 46 : Anouck et Alizée, épreuve 1, compréhension de texte, questions relatives à l'objectif 3

Durant le séminaire 2, Anouck¹⁵² explique : « *C'est la première fois que c'est un test décroché. D'habitude on évalue par rapport à des textes lus en classe. Cette fois, on a eu envie de partir sur un texte inconnu. C'est pratique : on le fait quand on veut* ». Il y a donc de l'inédit dans ses pratiques d'évaluation. Par ailleurs, elle justifie la présence des trois objectifs de la manière suivante : « *le 1 revient toujours, il est fondamental* », avoue ne pas trouver de sens à garder les objectifs du PER et privilégie leurs propres reformulations qui s'en inspirent, et enfin explique que ce raisonnement a modifié l'ordre des exercices. De son côté, durant le même séminaire, Alizée confirme les dires de sa collègue, à savoir que c'est un test décroché, et que l'objectif 1, fondamental, revient toujours.

5.1.2.2 Seconde évaluation créée pendant la formation

Six semaines se sont écoulées entre la discussion de la première et de la seconde évaluation. Cette dernière porte sur l'expression écrite (le texte explicatif). Elle se destine cette fois-ci à des élèves de 10 VP et sera mise en œuvre suite au séminaire. Elle est constituée de trois parties :

- Une page de garde sur laquelle figurent l'objectif du PER évalué (L1-32- Écrire des textes de genres différents adaptés aux situations d'énonciation...) et cinq de ses composantes choisies dans le plan d'études. Nous y trouvons également, pour le TS d'expression écrite, neuf éléments de progression des apprentissages de l'objectif en question, et pour un TA¹⁵³ de structuration, cinq éléments faisant référence au fonctionnement de la langue ;
- Une page de consignes d'écritures, ainsi que l'explicitation de la pondération concernant l'évaluation du fonctionnement de la langue ;
- Une page sur laquelle figure l'échelle, construite sur la base d'éléments de contenus.

Les éléments de discussion étant nombreux, nous les relevons en fonction de chaque partie.

¹⁵² Certains propos que nous relevons sont tenus par Alizée, avec qui Anouck a construit l'évaluation, mais étant donné qu'elles l'ont présentée et discutée ensemble, nous partons du principe qu'Anouck les partage, et que, de ce fait, ils peuvent être considérés comme siens.

¹⁵³ Les enseignantes de cet établissement parlent d'ETA (élément de travail assimilé) alors que la terminologie est le TA (travail assimilé). Pour des raisons d'homogénéité de discours, nous restons sur la seconde appellation, mais il faut considérer ces deux termes comme synonymes.

Objectif :

L1 32 — Écrire des textes de genres différents adaptés aux situations d'énonciation...

- ... en organisant ses idées, en personnalisant son message et en précisant sa pensée
- ... en analysant la situation d'énonciation et en s'y adaptant
- ... en organisant son texte en fonction des lois du genre (organiseurs, mise en page, temps verbaux)
- ... en respectant les règles de l'expression écrite (temporalité, orthographe, syntaxe, reprises, ...) et en utilisant un vocabulaire adapté à la situation
- ... en utilisant la dimension créative de la langue et sa propre créativité

Progressions des apprentissages requis :

Pour le TS d'expression écrite :

Identification des composantes de situations d'énonciation diverses : auteur/émetteur, public destinataire, visée
Élaboration des contenus : recherche des idées ou sélection des informations
Organisation du texte (planification du texte, choix ou respect du registre de langue, du système de temps, de la forme du texte, ...)
Choix du vocabulaire
Reprises pronominales et nominales
Mise en page structurée du texte : titre, sous-titres, paragraphes, illustrations, croquis, schémas, ...
Utilisation d'une structure explicative, des procédés explicatifs de la définition, de l'exemplification, de l'énumération, de la description, des organisateurs énumératifs, chronologiques, additifs, ...
Défense d'une thèse avec des arguments vraisemblables, illustrés par des exemples, en utilisant des marques de subjectivité et réfutée par une thèse adverse
Apport d'une conclusion pertinente exprimant un avis

Pour l'ETA de structuration :

Vérification et amélioration des temps verbaux
Respect de la syntaxe
Utilisation correcte de la ponctuation
Respect de l'orthographe lexicale et grammaticale des mots
Utilisation de moyens de références

Pour réviser :

- Relire la séquence « Expliquer – argumenter » p.238 à 259 LU, particulièrement le texte p.242 et les encadrés violets.
- Choisir un phénomène en lien avec le climat (cela peut être le même que pour votre exposé oral) et documentez-vous. Vous devez particulièrement préparer le vocabulaire des sciences de la Terre pour pouvoir utiliser les mots précis concernant votre phénomène. Un lexique très complet peut vous aider : <http://www.notre-planete.info/environnement/vocabulaire.php>

Figure 47 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), page de garde

Les éléments de discussion proposés par le formateur sous formes de questions permettant un débat sont ici les suivants :

- Une interrogation sur les caractéristiques du genre encyclopédique, puisque c'est le genre de texte que les élèves vont devoir produire ;
- Une remarque en rapport avec le fait que les éléments de progression des apprentissages du PER émanent de plusieurs genres (texte qui transmet des savoirs, texte qui argumente) ;
- Une question sur le choix des éléments de fonctionnement de la langue évalués – par ailleurs relevés comme intéressants - qui, formellement, ne correspondent à aucun objectif du PER.

La deuxième partie de l'épreuve concerne spécifiquement la tâche évaluative.

Vous allez rédiger un texte explicatif qui cherche à convaincre le public ado d'un magazine spécialisé de la nécessité d'agir en faveur du développement durable à propos d'une thématique précise sur laquelle vous avez fait antérieurement des recherches.

CONSIGNES D'ECRITURE :

1. Sur le modèle du texte p. 242 LU, vous allez écrire un court texte (100 - 150 mots) qui répondra aux questions suivantes :
 - Quelles sont les causes du problème décrit ?
 - Quelle est la thèse que vous défendez ?
 - Quel argument pourrait contredire votre thèse et comment réfutez-vous cet argument ?
 - Quelle est votre conclusion quant à ce problème/phénomène/... ?
2. Votre texte devra pouvoir être publié dans un journal pour ados, donc le vocabulaire scientifique devra être précis, mais l'ensemble compréhensible facilement.
3. Vous le présenterez dans une mise en page facilitant la lecture : propreté, paragraphes, sous-titres éventuels.
4. Vous trouverez un titre percutant.
5. Vous soignerez le vocabulaire, éviterez les répétitions et varierez les reprises.

Figure 48 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), consignes de production

Là, le formateur questionne le rapport entre les éléments de progression des apprentissages et les consignes : « De mon point de vue, les PA 1 et 3 ne se retrouvent pas dans vos consignes. Qu'en dites-vous ? ». Ensuite, il tente, sur la même page, de comprendre le système de

pondération et de notation élaboré pour l'évaluation du fonctionnement de la langue au travers d'un TA : « Ici, 6 points maximum, pour des notes allant de 1 à 6, sans demi-note ? », tout en questionnant le rapport entre les éléments évalués et les objectifs annoncés : « Et les répétitions, les reprises ? (j'essaie d'aligner ;-)) ».

<u>Technique de la langue : ETA</u>		<u>Note :</u>
On tiendra compte des 100 premiers mots pour l'ETA. ➤ N'oubliez pas d'indiquer le 100 ^e mot par un *		
Orthographe :	0 à 2 erreurs : 2 pts	3 à 4 erreurs : 1.5 pt
	5 à 6 erreurs : 1 pt	7 à 8 erreurs : 0.5 pt
	9 erreurs ou plus : 0	
Conjugaison :	0 à 1 erreur : 2 pts	2 à 3 erreurs : 1 pt
Syntaxe (phrases correctes, connecteurs utilisés correctement, ...) et ponctuation :	0 à 1 erreur : 2 pts	2 erreurs : 1.5 pt
	3 à 4 erreurs : 1 pt	5 erreurs : 0.5 pt
	6 erreurs ou plus : 0 pt	

Figure 49 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), pondération et notation du TA de fonctionnement de la langue

Puis, vient la troisième page, qui explicite les modalités de notation au travers de critères.

<p>Pour avoir la note correspondante, tous les éléments indiqués doivent être présents dans le texte !</p>	
6	<p>Doit comporter tous les éléments pour le 4, le 4.5, le 5 et le 5.5, mais, en plus :</p> <ul style="list-style-type: none"> - Le texte dans son ensemble est d'une grande cohérence. - Il est très original et ne contient aucune maladresse.
5.5	<p>Doit comporter tous les éléments pour le 4, le 4.5 et le 5 mais, en plus :</p> <ul style="list-style-type: none"> - le vocabulaire est varié, sans aucune maladresse. - Les arguments sont particulièrement intéressants et bien illustrés par des exemples. - Le contre-argument est réfuté de manière intéressante.
5	<p>Doit comporter tous les éléments pour le 4 et le 4.5, mais, en plus :</p> <ul style="list-style-type: none"> - le titre est intéressant. - Il y a des reprises utilisées correctement. - Il n'y a pas plus d'une répétition péjorant la qualité du texte. - Le vocabulaire est utilisé correctement, une légère maladresse est tolérée.
4.5	<p>Doit comporter tous les éléments pour le 4, mais, en plus :</p> <ul style="list-style-type: none"> - Présence d'une thèse, d'une thèse adverse et d'une conclusion vraisemblables et intéressantes. - Marques de subjectivité présentes. - Exemples et arguments intéressants.
4	<ul style="list-style-type: none"> - Présence d'une thèse, d'une thèse adverse et d'une conclusion vraisemblables. - Tous les éléments des consignes d'écriture sont présents, sans exception.
3.5	<ul style="list-style-type: none"> - Il manque un élément important du texte (argument, exemple, thèse, conclusion).
3	<ul style="list-style-type: none"> - Il manque des éléments des consignes d'écriture. - Le texte est trop court ou incohérent.
2.5	<ul style="list-style-type: none"> - Le texte est peu compréhensible, mal construit, pas clair, il présente de très importants défauts indiqués sur le texte.

Figure 50 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), échelle de notation

Là, le système de pondération et la construction de la note étant très novateurs, le formateur s'est centré sur des questions visant à demander aux enseignantes de justifier leurs pratiques, d'un point de vue de cohérence d'alignement curriculaire élargi. Pour cela, il leur pose trois

questions :

- La première est relative au poids de certains apprentissages dans le système élaboré : « Système très intéressant !! Cela signifie-t-il que certaines progressions d'apprentissage du PER, comme celle relative aux marques de subjectivité, sont pondérées plus fortement que d'autres ? » ;
- La deuxième réfère à la construction du seuil du suffisance (note 4) : « Comment êtes-vous arrivées à ce minimum requis en termes de contenu ? » ;
- La dernière concerne la formulation des critères : « Par rapport à la formulation de critères, parfois vous êtes précises (présence d'une thèse), parfois vous utilisez des adjectifs qui prêtent à interprétation (intéressant, original, incohérent). Il en va de même pour les noms : qu'est-ce qu'une maladresse en vocabulaire ? Comment gérez-vous ces différences ? ».

Notons que lors du séminaire 3, Anouck commence par dire que cette évaluation n'est pas satisfaisante : « *en fait, le titre devrait être expliquer pour argumenter* ». Elle veut ici expliciter la finalité de l'explication pour les élèves. Relevons également que les contenus évalués entre les deux épreuves varient énormément. Alizée, quant à elle, avalise le fait que le titre devrait être « *expliquer pour argumenter* ». Par ailleurs, elle explicite le processus de jugement qui lui a permis, avec sa collègue, d'abandonner les points au profit des critères et surtout d'une note référée à l'apprentissage, processus sur lequel nous revenons plus loin. Malgré cet état de fait, il est intéressant de constater que de nombreux changements sont intervenus d'une fois à l'autre au sein des pratiques évaluatives des deux enseignantes :

- Abandon de la traduction des objectifs du PER ;
- Homogénéisation du degré de complexité et du contenu des tâches évaluatives, notamment au niveau des consignes ;
- Systèmes de pondération différenciés (en points pour évaluer des éléments simples, en critères pour ce qui concerne des apprentissages complexes) ;
- Construction d'une note référée à l'apprentissage.

Certes, des éléments méritent une analyse plus fine. Raison pour laquelle nous allons maintenant vers notre triangulation de données.

5.1.2.3 Deuxième écrit réalisé pendant la formation

Anouck

Le texte d'Anouck fait environ deux pages (annexe Partie 5, 7.5). Elle en parle lors du séminaire 3, car elle y a consigné somme de réflexions en rapport avec la seconde évaluation qu'elle présente aux collègues. Ce texte a été rédigé en plusieurs étapes, qui suivent la chronologie de la création et de la mise en œuvre de cette évaluation, et qui donnent à voir les mouvements dans ses pratiques.

Tout d'abord, elle dit : « *Lors du premier cours, j'ai déjà compris que je devais mieux poser mes objectifs, aligner correctement jusqu'au bout et à partir de ce moment, je me suis sentie plus à l'aise dans ma conception de test* ». Elle montre ici potentiellement un autre rapport au PER, comme nous pouvons le supposer dans sa seconde épreuve.

Puis, elle met des mots sur un des phénomènes dérangeants qu'elle ne pouvait nommer avant la formation, et qu'elle pense désormais être capable de dépasser : « *Puis lors du deuxième cours m'est apparue l'évidence de ce que je ne parvenais pas à formuler, le je-ne-sais-quoi qui, finalement, me dérangeait tout le temps : la compensation. Pourquoi j'avais parfois l'impression que des élèves n'avaient rien compris au sujet, lorsque, comme à chaque fois, je corrigeais les tests par questions (toutes les unes, les deux, ...), mais qu'au final, une fois les points additionnés, ils avaient une bonne note. L'injustice était peut-être là...* ». Là, nous retrouvons cette recherche de justesse en évaluation qui était au cœur de ses priorités au départ. Pratiquement, elle voit des possibles, différents en fonction des contenus : « *Pour la compréhension et la grammaire, la réponse m'a été donnée : tester un seul critère. Accepter de ne pas tout tester finalement* », « *Pour l'expression écrite, j'ai voulu me lancer dans une grille 100% qualitative, n'offrant pas de compensation possible. Après quelques jours de macération du projet et une relecture par Alizée, prête à me suivre, cela prend forme finalement assez facilement, comme une évidence...* », en référence à la seconde évaluation.

Ce revirement, selon ses dires, la rend « *euphorique* ». Elle semble progresser énormément dans ses pratiques, dans son questionnement. Certes, son barème ne la satisfait pas après usage, et elle questionne encore et encore la pertinence de ses tâches évaluatives. Mais les notes semblent correspondre aux performances des élèves. Toutefois, le doute, émanant de nouvelles pratiques possiblement encore fragiles, est encore présent : « *Est-ce uniquement une question*

d'habitude ? » se demande-t-elle, faisant référence au poids des pratiques usitées depuis trop longtemps. Quoi qu'il en soit, le mouvement est en marche : *« Je suis donc à l'aise avec les petites révolutions que je suis en train de mener et je compte creuser la question »*, malgré une interrogation sur la possible transposition de telles démarches à l'évaluation du fonctionnement de la langue, encore.

Alizée

Le texte d'Alizée, quant à lui, est un des plus longs récoltés : trois pages (annexe Partie 5, 7.2). Il a été rédigé en cinq étapes, comme un journal de bord, qui suivent la chronologie de la formation. Il commence quatre jours après le séminaire 1 et se termine cinq jours avant le dernier, sur une période de deux mois et demis. Alizée y relate diverses tentatives de mise en œuvre des contenus et réflexions partagées en formation.

Le 16 février, elle dit être confortée dans son rapport au PER suite à la découverte du modèle théorique à visée compréhensive : *« j'ai trouvé le début de la théorie plaisant et logique. Il me plaît bien d'aligner les objectifs avec leurs composantes et leurs progressions d'apprentissage. Ainsi que de faire des choix qui correspondent à mon enseignement. Cela ressemble à ce que je faisais déjà et c'est toujours agréable de se sentir conforté dans sa pratique et même de trouver une justification théorique à nos choix pratiques »*. C'est un élément étonnant, Alizée ayant affirmé que les objectifs faisant référence pour elle étaient réinterprétés a posteriori de ceux du PER. Ensuite, elle dit devoir être vigilante à aligner tâches et objectifs : *« Ensuite le fait de vérifier que nos exercices correspondent à ces objectifs choisis me semble aussi assez évident. Quoique je me demande si c'est vraiment toujours le cas et que je me promets de vérifier dans mes prochains tests si cela l'est vraiment »*. Puis, elle exprime deux questions fortes, qui se retrouvent déjà dans ses interrogations de départ : comment mettre les points ? Et comment choisir ses échelles : *« la question de l'échelle est celle qui me remet le plus en question... »*.

Deux jours plus tard, elle reprend la plume. Elle explicite sa manière d'élaborer avec Anouck la première évaluation précédemment présentée. D'abord, traduire les objectifs et les mettre en rapport avec des questions. Ensuite, la mise de points au regard des contenus, jusqu'à l'élaboration du 4 : *« je réfléchis à plusieurs possibilités pour mettre les points qui pourront dépendre aussi du poids que nous souhaitons attribuer à chaque. Puis je me demande aussi quelles sont les questions qui devraient être réussies par les élèves pour obtenir le 4 »*.

Suite à la rencontre avec Anouck, elle rédige la suite des réflexions le 7 mars. Là, elle explicite leur manière de résoudre la question : « *Nous étions assez d'accord sur la manière de mettre les points. Nous avons ensuite regardé chaque exercice et estimé ce que les élèves devaient atteindre pour avoir le 4. Le total auquel nous sommes arrivées est 20 points sur les 30 du test ce qui peut paraître assez large. Mais cela correspond bien à la difficulté du test, qui reprend tout ce que nous avons vu en compréhension de texte cette année. Nous avons ensuite réfléchi aux autres notes. Le 6 et le 5,5 nous paraissent devoir être assez serrés : c'est l'excellence, les élèves qui ont été capables de répondre correctement à presque toutes les questions. Puis nous répartissons les points équitablement entre le 5, le 4,5 et le 4. Pour en-dessous, nous décidons d'élargir progressivement* ».

La suite de son texte est écrite suite au deuxième séminaire. Elle se dit étonnée de voir combien les pratiques ont déjà évolué. Puis, elle décline les étapes de construction de la seconde évaluation, également présentée plus haut, en mettant en évidence les prises de consciences et mouvements opérés depuis le départ, à savoir :

- Abandon des points au profit d'une évaluation critérielle ;
- Exploitation du PER et sélection des objectifs clés ;
- Centration sur les contenus ;
- Séparation du fonctionnement de la langue et de la production écrite dans l'évaluation ;
- Reformulation des consignes en références aux attendus ;
- Affinage des critères de manière à permettre une notation référée aux apprentissages.

Le saut est à ses dires très important, voire « *effrayant* ». Une question la déstabilise : « *que ferons-nous si un élève oublie juste un élément de consigne (par exemple le titre) mais que son texte contient des éléments qui lui vaudrait un 4,5 ou un 5 ?* ».

Dix jours plus tard, elle tire un bilan très positif de l'usage de la grille. Pas de catastrophe, et un soulagement : « *on ne se prend plus la tête avec le choix du nombre de points et qu'il n'y a plus besoin de faire d'échelle* ».

Des mouvements semblent donc bel et bien être en marche, en référence à son questionnement de départ.

5.1.3 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives d'Anouck à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *pendant la formation*

Sont prises en compte ici trois des quatre sources de données particulièrement significatives durant la formation :

- Les éléments de discussion proposés par le formateur (annexe Partie 5, 9). Mentionnés plus haut, ils se réfèrent tous au modèle théorique de l'alignement curriculaire élargi. Ils sont reproduits dans le tableau ;
- Les éléments débattus en séminaires de formation 2 et 3, retranscrits depuis les enregistrements ;
- Le deuxième écrit documenté précédemment.

Le tableau ci-dessous rend compte de la triangulation de ces données référée à nos catégories conceptualisantes.

Ce tableau nous permet notamment de relever qu'un seul élément de discussion a mené à des débats lors du séminaire 3, et cela, malgré les changements importants amenés par Anouck et Alizée. Nous tentons d'en expliciter les raisons plus loin. Les éléments étant ici différents, le premier tableau et ses commentaires concerne Anouck, le second Alizée.

Tableau 14 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Anouck durant la formation

Données Catégories	Evaluation 1 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Anouck débattus en séminaire 2 de formation	Evaluation 2 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Anouck débattus en séminaire 3 de formation	Ecrit 2
Rapport au référentiel	Questionnement du processus de reformulation des objectifs du PER	Différences entre les objectifs du PER et ceux reformulés : « <i>la division est plus claire, on voulait montrer la progression</i> »	Les éléments de progression des apprentissages du PER émanent de plusieurs genres		
	Quelle cohérence entre les tâches et les objectifs reformulés vs ceux du PER ?	Le L1-31 regroupe les objectifs 1 et 3	Questionnement du rapport entre les éléments de progression des apprentissages et les consignes		

	Quelle pertinence d'évaluer le L1-36, ou de le séparer du L1-31 ?				
Expression de l'objet	Les caractéristiques principales du genre de texte abordé sont-elles prises en compte ?	« On tient à ce que le vocabulaire soit au service de la compréhension » : quel est son statut dans l'évaluation ?	Interrogation sur les caractéristiques du genre encyclopédique	Elaborer des grilles de critères et une notation en référence à des unités d'apprentissage semble acquis en expression écrite. Mais qu'en est-il pour le fonctionnement de la langue ?	Elaborer des grilles de critères et une notation en référence à des unités d'apprentissage semble acquis en expression écrite. Mais qu'en est-il pour l'orthographe ?
	« Il y a beaucoup d'éléments. Sont-ils tous évalués ? »	Est-il question de compréhension ou d'analyse, comme le stipule le L1-31 ? Le PER ne parle pas de compréhension, comment gérer ce manque ?	Justification du choix des éléments de fonctionnement de la langue évalués		
	Evalue-t-on de la compréhension ou de l'analyse ?				
Elaboration des tâches évaluatives		Les questions travaillent-elles et/ou évaluent-elles la compréhension ?		Le décalage entre certaines consignes et les objectifs est relevé et accepté	Les items sont à revoir, encore et toujours
Pratiques de pondération	Questionnement du nombre de points attribués à chaque objectif	Un objectif pondéré à hauteur de 3,5 pts, est-ce pertinent ? Oui, car le 1 est incontournable	Questionnement du rapport entre les éléments évalués et les objectifs annoncés	Questionnement du degré de précision et de transparence des critères	
	Comment sont attribués les points ?		Questionnement du poids de certains apprentissages dans le système élaboré		

Pratiques de notation	« Pourquoi les échelons s'agrandissent-ils au fur et à mesure que l'on « descend » ? Comment gérer les différents objectifs évalués avec une somme de points ? »	Est-ce envisageable de faire deux notes ? Une de vocabulaire au travers d'un TA, l'autre en compréhension en TS ?		Seuil du suffisance (note 4) : « Comment êtes-vous arrivées à ce minimum requis en termes de contenu ? »	Le barème, après utilisation, ne donne pas satisfaction
-----------------------	--	---	--	--	---

Tableau 15 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Alizée durant la formation

<div>Données</div> <div>Catégories</div>	<i>Evaluation 1 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Alizée débattus en séminaire 2 de formation</i>	<i>Evaluation 2 : éléments de discussion proposés par le formateur</i>	<i>Eléments acquis et/ou de discussion pour Alizée débattus en séminaire 3 de formation</i>	<i>Ecrit 2¹⁵⁴</i>
Rapport au référentiel	Questionnement du processus de reformulation des objectifs du PER	Différences entre les objectifs du PER et ceux reformulés : « <i>la division est plus claire, on voulait montrer la progression</i> »	Les éléments de progression des apprentissages du PER émanent de plusieurs genres		Prise en compte des objectifs du PER sans leur reformulation
	Quelle cohérence entre les tâches et les objectifs reformulés vs ceux du PER ?	Le L1-31 regroupe les objectifs 1 et 3	Questionnement du rapport entre les éléments de progression des apprentissages et les consignes		
	Quelle pertinence d'évaluer le L1- 36, ou de le séparer du L1- 31 ?				
Expression de l'objet	Les caractéristiques principales du genre de texte abordé sont- elles prises en compte ?	« <i>On tient à ce que le vocabulaire soit au service de la compréhension</i> » : quel est son statut dans l'évaluation ?	Interrogation sur les caractéristiques du genre encyclopédique	Elaborer des grilles de critères et une notation en référence à des unités d'apprentissage semble acquis	Centration sur les contenus

¹⁵⁴ Nous ne prenons ici en considération que les éléments les plus récents et significatifs du parcours décrit par Alizée dans son texte.

	« Il y a beaucoup d'éléments. Sont-ils tous évalués ? »	Est-il question de compréhension ou d'analyse, comme le stipule le L1-31 ? Le PER ne parle pas de compréhension, comment gérer ce manque ?	Justification du choix des éléments de fonctionnement de la langue évalués	en expression écrite. Mais qu'en est-il pour le fonctionnement de la langue ?	Séparation du fonctionnement de la langue et de la production écrite dans l'évaluation
	Evalue-t-on de la compréhension ou de l'analyse ?				
Elaboration des tâches évaluatives		Les questions travaillent-elles et/ou évaluent-elles la compréhension ?		Le décalage entre certaines consignes et les objectifs est relevé et accepté	Attention accrue à aligner tâches et objectifs
					Reformulation des consignes en références aux attendus
Pratiques de pondération	Questionnement du nombre de points attribués à chaque objectif	Un objectif pondéré à hauteur de 3,5 pts, est-ce pertinent ? Oui, car le 1 est incontournable	Questionnement du rapport entre les éléments évalués et les objectifs annoncés	Questionnement du degré de précision et de transparence des critères	Abandon des points au profit d'une évaluation critérielle
	Comment sont attribués les points ?		Questionnement du poids de certains apprentissages dans le système élaboré		Affinage des critères de manière à permettre une notation référée aux apprentissages
Pratiques de notation	« Pourquoi les échelons s'agrandissent-ils au fur et à mesure que l'on « descend » ? Comment gérer les différents objectifs évalués avec une somme de points ? »	Est-ce envisageable de faire deux notes ? Une de vocabulaire au travers d'un TA, l'autre en compréhension en TS ?		Seuil du suffisance (note 4) : « Comment êtes-vous arrivées à ce minimum requis en termes de contenu ? »	

Cette analyse au travers des catégories conceptualisantes donne à voir sur quelles dimensions de leur pratique Anouck et Alizée, avec l'aide de leurs pairs et du formateur, portent leurs réflexions. Nous l'avons montré, des changements importants sont opérés durant la formation. Tentons ici de les reprendre, en les mettant en perspective avec d'autres traces récoltées. Pour Anouck :

- Le PER trouve un sens praxéologique puisqu'il peut désormais être exploité de manière critique. Certes, des apprentissages restent formulés en référence aux moyens utilisés (comme le TA de l'épreuve 2), mais Anouck parvient là à justifier ses choix à ce niveau pour le TS : *« le choix des progressions est lié à l'explication, avec une idée d'argumentation, mais sans que ce soit le but recherché »* ;
- Son analyse des items et/ou des situations d'évaluation se rapporte à leur cohérence avec les objectifs ;
- Le système de pondération par des critères semble combler ses doutes : *« Je ne veux plus travailler qu'avec des grilles comme ça »* dira-t-elle à ses collègues lors du séminaire 3. Au travers de cet enthousiasme, elle met en évidence l'aisance avec laquelle elle peut justifier ses choix, notamment parce qu'elle perçoit l'inscription de ses démarches dans des apprentissages analysés et des pratiques de notation transparentes et exigeantes ;
- Enfin, elle réalise que sa nouvelle manière de construire la note en référence à des unités d'apprentissage lui permet d'explicitier ses attentes et dès lors de tendre vers cette justesse à laquelle elle tient tellement. Le fait qu'elle pratiquait déjà avant la formation un calcul du seuil de suffisance en points n'est hypothétiquement pas étranger à ce mouvement significatif dans les pratiques.

Pour Alizée :

- Le PER paraît trouver un sens praxéologique puisqu'il peut désormais être exploité de manière critique. Certes, des apprentissages restent formulés en référence aux moyens utilisés (comme le TA de l'épreuve 2), mais Alizée abonde dans la justification de leurs choix à ce niveau pour le TS ;
- Les tâches se rapportent à leur cohérence avec les objectifs, c'est en tous cas un point sur lequel elle porte désormais son attention ;
- Le système de pondération par des critères semble lui donner satisfaction : *« La grille pour l'expression écrite est facile à utiliser. On liste d'abord les critères de base pour le 4, puis, s'ils y sont tous, on remonte progressivement pour voir ce qu'ils ont fait de*

mieux (ou pas) » écrit-elle. Cet enthousiasme se traduit également par des aspects pratiques, comme la communication aux élèves et la correction : « Ce qui est certain c'est que la correction est vraiment très rapide (30 minutes pour les 3), que les élèves ne semblent pas avoir eu de problème à comprendre cette manière de fonctionner, qu'on ne se prend plus la tête avec le choix du nombre de points et qu'il n'y a plus besoin de faire d'échelle. On gagne donc beaucoup de temps » ;

- Elle donne davantage d'importance aux contenus, via sa réflexion sur les critères ;
- Enfin, elle réalise que sa nouvelle manière de construire la note en référence à des unités d'apprentissage lui permet d'explicitier ses attentes et dès lors de tendre vers cette justesse à laquelle elle tient tellement.

Deux questions restent toutefois ouvertes. La première vient du duo : Comment transposer les démarches évaluatives qualitatives réalisées à des objets relatifs au fonctionnement de la langue ? La seconde, en revanche, émane de nos données et n'a pas été traitée par l'intéressée : Qu'en est-il du rapport de cohérence et surtout de son explicitation entre les tâches, les critères et les descripteurs des échelons pour établir la note ? Autant d'éléments qu'il nous faudra traiter à l'aide de nos dernières données récoltées en fin de formation.

À la lumière de ces éléments, nous sommes en mesure de formuler quelques hypothèses interprétatives traduisant des mouvements dans les pratiques évaluatives des deux enseignantes, et qui nous permettent de voir dans quelle mesure elle parvient à travailler leur cohérence du point de vue du modèle théorique :

- Le PER semble être une référence dont les objets de savoir peuvent, dans certaines conditions, « circuler » jusque dans une notation référée à l'apprentissage ;
- Les tâches sont potentiellement inscrites dans des objets de savoir sur lesquels Anouck a un regard critique, et donnent à voir les éléments de l'apprentissages importants ;
- La pondération peut s'outiller de manière complexe lorsque les apprentissages évalués le méritent, mais aussi prendre une forme plus traditionnelle lorsqu'ils sont taxonomiquement bas. Là, Alizée explique : « Nous¹⁵⁵ ferons un TA de structuration séparé ce qui permet de n'évaluer dans le TS que les éléments de l'objectifs L1 32 que nous avons choisis » ;

¹⁵⁵ Le « Nous » est ici pour nous emblématique de l'étroite collaboration entre les deux enseignantes.

- La notation, dans certaines situations, peut être en cohérence avec un système de pondération critérié, lui-même référé à des tâches complexes évaluant des objectifs ambitieux.

En revanche, à la lumière de nos données, il nous semble difficile de dégager des tensions ou difficultés à dépasser des ruptures d'alignement curriculaire élargi. Nous y reviendrons dans notre partie de discussion-conclusion.

Après la formation

5.1.4 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives d'Anouck à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *à la fin de la formation*

Anouck

Anouck a produit son troisième écrit la veille du dernier séminaire de formation. Le bilan ainsi produit met en exergue les trois points positifs suivants :

- Limiter la quantité de matière évaluée ;
- Eviter la compensation par l'usage de grilles critériées et favoriser tant que possible la construction de notes en référence à des unités d'apprentissage ;
- Délimiter le seuil du 4 en référence à des contenus importants.

Le tableau ci-dessous rend compte des éléments significatifs de deux sources de données relativement aux points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre : le dernier écrit et le second entretien. Ils sont directement mis en rapport avec les catégories conceptualisantes, de manière à rester synthétiques et centrés sur nos questions de recherche. Certes, des redondances peuvent apparaître, mais elles doivent à nos yeux être comprises et interprétées comme des signes d'un processus complexe d'appropriation de divers éléments par Anouck. Nous commentons ces données plus bas en les mettant en dialogue avec les autres données.

Tableau 16 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Anouck à la fin de la formation

<i>Données</i>	<i>Ecrit 3</i>	<i>Entretien 2</i>
Catégories		
Rapport au référentiel	« J'ai pu légitimer ce que je pensais déjà : si on doit enseigner tous les éléments du PER, on ne doit pas tout évaluer »	PER davantage pris en compte, même si « j'ai pas besoin de passer par la formulation exacte »
		« J'ai l'impression d'évaluer correctement, donc du coup, j'ai pas besoin de le mettre [le PER] »
		« ça va tellement dans tous les sens, et puis c'est tellement large, y a tellement de compétences qui sont liées entre elles, que de formuler un objectif pour un exercice, c'est impossible, quasiment »
Expression de l'objet	« Je vais à l'avenir limiter la variété de la matière »	« Qu'est-ce qui, pour nous, est important ? Qu'est-ce qu'on veut voir dans le test ? Donc c'est une manière de faire ressortir... le fondamental. Enfin, je veux dire, ça nous remet directement au cœur de la matière »
		Ne pas tester beaucoup de choses
Elaboration des tâches évaluatives		
Pratiques de pondération	« J'ai mis en place deux importantes modifications : imaginer une grille qualitative et dissocier la partie structuration et textuelle »	Retour à des critères moins précis
	« C'est paradoxal car en regardant ma grille, elle donne l'impression d'être beaucoup moins objective, mais je l'assume pleinement »	« J'arrive à voir ce qui va pas dans la compensation » Sélection précise des critères clés
	3 changements : « limiter le nombre de critères, chercher à éviter lorsque cela me semble important la possibilité de réussir par compensation et utiliser des grilles qualitatives »	« Les points, c'est pas la matière » ; « on perd beaucoup de cohérence avec ça » Moins de critères : la grille qualitative, « une révolution »
Pratiques de notation	« Je partirai dorénavant des minima attendus pour fixer mon 4 et plus d'un pourcentage »	Dans une perspective qualitative, difficulté à passer d'une note à l'autre (différence entre un 4 et un 4,5 ?)
		« la grosse question que j'ai encore, c'est celle de l'échelle. [...] je place très bien mon 4, [...] je me sens légitimée dans ma manière de placer le 4. Je le fais facilement, après il reste le problème d'en dessus et d'en dessous quoi ». Les grilles qualitatives règlent ces questions.
		« Et puis là, maintenant je me rends compte que ça [le barème standardisé] va pas avec le reste en fait »

		La linéarité des échelles est encore un dilemme « j'évite la rupture avec l'échelle après »
--	--	--

Ce tableau nous permet de formuler des hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi :

- L'expertise des apprentissages et des objets de savoir qui semble en développement permet un rapport au PER plus critique, notamment au niveau des choix des objectifs à travailler et à évaluer. À ce titre, certains renoncements deviennent possibles, ce qui est nouveau pour Anouck ;
- L'objet de savoir paraît prendre une place centrale dans les raisonnements, qu'il se situe au niveau du « quoi évaluer ? » ou de la notation ;
- Le « retour » à des grilles de critères moins analytiques est argumenté ;
- Une notation référée à l'apprentissage semble se stabiliser, sauf sur certains contenus, comme le fonctionnement de la langue.

S'en dégagent également deux tensions. Au regard de nos données, nous faisons l'hypothèse qu'elles ne traduisent toutefois pas des difficultés à dépasser des ruptures d'alignement curriculaire élargi. En effet, elles semblent permettre des compromis qui n'invalident pas cette cohérence recherchée dans l'évaluation :

- Le respect d'une linéarité des échelles : Anouck affirme qu'elle est encore en dilemme sur la linéarité à construire avec des grilles critérielles. Cette réflexion interroge le statut des critères et leur contenu. Nous y reviendrons dans la partie suivante ;
- La reformulation des objectifs du PER semble rester de mise. Il faudra alors problématiser l'écart potentiel entre la prescription et l'interprétation que ce travail peut représenter, ainsi que le statut du PER comme référentiel d'évaluation. Nous y reviendrons également dans notre conclusion.

Ces derniers constats nous autorisent à faire l'hypothèse que l'alignement curriculaire élargi apparaît ici comme une lunette analytique compréhensive très aidante. Anouck, dans son second entretien, y fait d'ailleurs souvent référence : « *Enfin, je veux dire, je pense que j'ai vraiment cette conscience maintenant de..., alors, je le faisais déjà au début, mais maintenant j'ai conscience que ça [l'alignement curriculaire] va jusqu'à l'échelle. Enfin jusqu'à la note* » (87). Pour elle, ce modèle lui a permis de remettre au centre les contenus : « *Parce qu'à partir*

du moment où on choisit nos objectifs, enfin qu'on décide clairement ce qu'on veut évaluer comme matière, je vois pas à quel moment ça pourrait nous poser problème d'aligner. Maintenant, l'enjeu, pour moi, il est aussi au début, enfin, justement, je l'ai dit plusieurs fois, au choix de la matière » (90). Des progrès, relevés dans nos données, sont explicités à ce niveau : « avant, je pense que je choisisais trop de matière. Du coup, j'arrivais pas à lui donner de la valeur jusqu'au bout. C'est ça, c'était un petit peu dilué, et puis la matière du début elle apparaissait plus du tout dans la note, y avait plus cet enjeu-là. C'était aligné en partie. Mais, je veux dire, il avait plus de valeur en fin de parcours. Ou très peu » (91-93).

Alizée

Alizée a produit son troisième écrit dix jours après le séminaire 3. Le texte, assez bref, fait un bilan qui met en évidence les points suivants, nuancés et critiques parfois :

- Le PER peut être envisagé comme un référentiel dont les contenus pourraient être visibles jusque dans la notation, même si l'opération, pour Alizée, est délicate : *« le PER n'est pas forcément complet quant aux objectifs que je souhaite évaluer. Ou alors, il est toujours difficile d'être sûr que l'échelle choisie reflète bien l'acquisition de l'objectif choisi »* ;
- L'évaluation fait partie du processus d'enseignement et d'apprentissage : *« L'alignement n'est plus seulement dans la construction de l'évaluation mais déjà dans les contenus enseignés »* ;
- Des changements profonds au niveau des pratiques de pondération et de notation ont pu apparaître : *« Grâce à la collaboration avec Anouck, nous avons réussi à laisser tomber les points et les échelles dans les tests d'expression écrite et à nous concentrer sur des critères de qualité »*. Ainsi, ses notes semblent avoir gagné en cohérence : *« Les notes correspondent à nos impressions de départ et nous nous sentons cohérentes »* ;
- L'expertise sur les contenus paraît être plus précise : *« J'aime bien aussi l'idée de séparer l'évaluation des contenus de celle de la forme »*. Il reste toutefois des questions relativement à l'évaluation de l'oral (compréhension et production), mais Alizée semble confiante en ses capacités à transposer une partie des outils qu'elle s'est appropriée.

C'est enfin un sentiment de confiance en ses pratiques et ses choix qui s'est consolidé chez Alizée : *« Plus généralement, après ce cours, je me sens plus libre. Vous nous avez redonné la confiance en notre capacité de spécialistes de faire des choix »*. Rappelons que le tableau ci-dessous rend compte des éléments significatifs de deux sources de données relativement aux

points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre : le dernier écrit et le second entretien.

Tableau 17 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives d'Alizée à la fin de la formation

<i>Données</i>	<i>Ecrit 3</i>	<i>Entretien 2</i>
Catégories		
Rapport au référentiel	Exploitation toujours délicate, notamment jusqu'à la notation	« je suis plus au clair avec ça [objectif-composantes-progressions]. Je suis plus... plus précise »
		« reformulations en utilisant le vocabulaire que j'ai utilisé quand j'ai enseigné »
		Liberté de formuler également des apprentissages qui ne figurent pas dans le PER : « y a quand-même certains sujets, où y a pas ce qu'il faut dedans. Voilà, il manque des choses, pour moi, dans ce PER »
Expression de l'objet	Centration sur des éléments de contenu	Clarification de la différence entre comprendre et analyser
		Le fonctionnement de la langue et les critères : quelle compatibilité ?
		Les contenus influencent les grilles de critères
		La pondération sur les contenus : « plus vite corrigé »
		« les résultats reflètent est-ce qu'ils savent ou pas ? »
		« On a évacué la structuration des grilles »
		autre agencement des objets dans les exercices
Elaboration des tâches évaluatives		Réflexion sur le rapport questions-objectifs
		« éviter de tout le temps faire des exercices sur la même habileté »
		« combien je veux dans chaque exercice pour que ça soit atteint ? »
		La taxonomie donne des idées d'exercices
		Consignes précises, concises
Pratiques de pondération	Exploitation de critères sur des apprentissages de production écrite	« le point représente un petit bout de progression »
		« choisir quels sont les critères indispensables, et quels sont ceux que je peux laisser tomber »
		Grilles simplifiées
		Concentration sur le contenu pour pondérer
		« on se demande quel contenu on évalue et ensuite combien ça vaut de points (nouveau) »

		« le plus complexe, c'était entre les points et l'échelle, il a fallu réfléchir, ben, comment faire pour aligner ça ? »
Pratiques de notation	Les notes ont gagné en cohérence	Pas besoin de marquer l'insuffisance en dessous de 3
		« raisonnement : il manque un de mes critères, il manque un de mes objectifs, c'est insuffisant »
		Autre manière de fixer le 4, en référence au contenu
		« que la note puisse être un reflet de l'acquisition d'un objectif »

Ici, les hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi :

- Comme chez sa collègue, l'expertise des apprentissages et des objets de savoir qui semble en développement permet un rapport au PER plus critique, notamment au niveau des choix des objectifs à travailler et à évaluer. À ce titre, certains ajouts deviennent possibles, ce qui est nouveau pour Alizée ;
- L'objet de savoir paraît prendre ici aussi une place centrale dans les raisonnements, qu'il se situe au niveau du « quoi évaluer ? », dans la création des critères ou dans la notation ;
- Les tâches sont pensées en référence à une taxonomie et aux objectifs qu'elles évaluent ;
- Le « retour » à des grilles de critères moins analytiques est également argumenté et assumé, même si leur usage sur le fonctionnement de la langue ou l'oral demande à être expérimenté ;
- Une notation référée à l'apprentissage semble se stabiliser, sauf sur certains contenus, comme le fonctionnement de la langue.

S'en dégage également une tension. Elle concerne ce travail extrêmement complexe de recherche de cohérence entre la pondération et la notation, qu'Alizée espère pouvoir mener à bien. Toutefois, au regard de nos données, nous faisons l'hypothèse qu'elle ne traduit pas des difficultés à dépasser des ruptures d'alignement curriculaire élargi. En effet, elle permet des compromis qui n'invalident pas cette cohérence recherchée dans l'évaluation : la reformulation des objectifs du PER semble rester de mise. Il faudra alors problématiser l'écart potentiel entre la prescription et l'interprétation que ce travail peut représenter, ainsi que le statut du PER comme référentiel d'évaluation. Nous y reviendrons également dans notre conclusion.

Ces derniers constats nous permettent de faire l'hypothèse que l'alignement curriculaire élargi

apparaît ici comme une lunette analytique compréhensive très aidante. Alizée, dans son second entretien, y fait d'ailleurs souvent référence, ainsi que son dernier écrit : *« la théorie de l'alignement... ouais. Ça je trouve qu'elle est très utile. Et même... Moi je suis allée regarder Biggs, là, même d'aligner avec l'enseignement, enfin ça, ça me parle beaucoup, quoi »* (entretien 2, 22-23) ; *« ça m'a rendue consciente de ce que je faisais ou pas, mais... Au départ, je me suis dit « mais c'est tout simple », enfin... et puis après, je me suis dit « ah mais non, c'est pas si facile » (rires) »* (entretien 2, 251). Pour elle, ce modèle lui a permis de remettre au centre le lien entre évaluation et enseignement : *« Mais ce qui serait intéressant, ça serait d'aller, ouais, d'aller plus loin, ça serait de commencer l'alignement par l'enseignement »* (entretien 2, 266), mais également l'importance des contenus : *« L'alignement n'est plus seulement dans la construction de l'évaluation mais déjà dans les contenus enseignés »*.

Tout au long de la recherche : Anouck

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse qu'Anouck est davantage outillée pour éviter des désalignements entre toutes les composantes du modèle théorique, lorsque les contenus s'y prêtent. À ce titre, elle ajoute encore que la taxonomie a été un outil très aidant pour penser cette cohérence en évaluation sommative : *« je pense que ça [la taxonomie] peut m'aider quand je construis mon test, en me disant, voilà. Mais même de reprendre la feuille. Parce que finalement c'est une aide énorme. C'est un peu, ça me donne des étapes quoi [...] je pense que ça me permettra de me dire « ok, maintenant je vais faire une compréhension de texte, donc c'est quoi « comprendre » ? Ok, c'est ça, ça, ça. Alors par rapport à mon texte, comment je vais, quelles questions je vais poser pour être sûre qu'ils ont compris ? » [...] Donc finalement, c'est un gain de temps de se dire « mon objectif, c'est comprendre un texte. C'est tout, y en a pas d'autre. Et je regarde quelle question je dois poser pour vérifier s'ils ont compris ou non le texte »* (100-101-102).

Il lui reste maintenant à explorer d'autres contenus dans une perspective qualitative, comme l'évaluation du fonctionnement de la langue ou la production de l'oral. Mais Anouck semble en confiance, forte des outils dont elle dispose désormais.

Nous pouvons dès lors appuyer nos hypothèses interprétatives par les divers changements importants en œuvre dans les pratiques évaluatives sommatives d'Anouck décrits plus haut, que nous considérons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi.

Tout au long de la recherche : Alizée

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse qu'Alizée est davantage outillée pour éviter des désalignements entre toutes les composantes du modèle théorique, lorsque les contenus s'y prêtent. À ce titre, elle ajoute encore que la taxonomie a été un outil très aidant pour penser cette cohérence en évaluation sommative, du point de vue des habiletés cognitives : « *La taxonomie d'Anderson et Krathwohl m'a été utile pour différencier la compréhension de l'analyse* ». Mais aussi au niveau des tâches : « *cette taxonomie-là, elle donne des idées, aussi, d'exercices. On se dit « ah, mais si je veux évaluer juste ça, alors qu'est-ce que je pourrais faire comme exercice... ».* Au lieu d'évaluer juste des contenus, on peut évaluer des... des manières de réfléchir, dans le cerveau » (entretien 2, 278). Toutefois, elle l'a trouvée moins évidente à exploiter comme un outil compréhensif.

Il lui reste maintenant à explorer d'autres contenus dans une perspective qualitative, comme l'évaluation du fonctionnement de la langue ou la production de l'oral. Mais Alizée semble en confiance, forte des outils dont elle dispose désormais.

Nous pouvons dès lors appuyer nos hypothèses interprétatives par les divers changements importants en œuvre dans les pratiques évaluatives sommatives d'Alizée décrits plus haut, que nous considérerons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi.

Mais qu'en-est-il maintenant du point de vue du développement potentiel des compétences évaluatives d'Anouck et d'Alizée, au regard de ces mouvements de pratique contrastés ?

5.2 Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez les deux enseignantes

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Anouck et Alizée au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe. Pour répondre à cette question, nous mettrons en correspondance les apports déjà dégagés à l'aide des catégories conceptualisantes avec les compétences construites pour notre étude dont le développement est présenté dans la partie méthodologique.

Il nous semble ici important de reprendre chacune des six compétences et d'en justifier le possible développement. Pour éviter de charger le texte, nous rappelons qu'elles sont décrites dans le portrait de Fernand qui ouvre ce chapitre. Aussi, en nous appuyant sur nos résultats précédents, nous pouvons documenter quelles compétences Anouck et Alizée semblent avoir développé :

Tableau 18 : Indices de développement potentiel des compétences évaluatives d'Anouck

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Oui, référé aux contenus, reformulations à l'aide d'une taxonomie comme outil analytique	Oui, en rapport avec les objectifs dans une perspective analytique			
B	Oui, traduisent des objets clés de la discipline présents dans le PER	Oui, traduisent des objets clés de la discipline	Oui, le plus souvent complexes du point de vue des habiletés cognitives et des contenus (compréhension)		
C	Oui, les critères sont formulés en référence aux contenus issus du PER	Oui, les critères sont pensés depuis les contenus		Oui, approches critériées sur des apprentissages complexes ¹⁵⁶	
D	Oui, la note tend à avoir un rapport avec les contenus issus du PER	Oui, la note rend compte d'éléments de contenus	Oui, la note traduit la complexité des objets présents dans les tâches	Oui, la note est issue d'un raisonnement référé aux unités d'apprentissages pour le calcul du seuil du 4, et dans certaines situations, pour toutes les notes	Oui, la note n'est plus construite sur une logique de compensation
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures évitables
F	Oui, rapport distancé	Oui, choix assumés	Oui, choix assumés référés à une taxonomie	Oui, le système critérié est considéré comme	Oui, les directives en la matière sont critiquées, voire refusées

¹⁵⁶ Nos données nous permettent de mettre un « oui » ici, même si des doutes subsistent chez Anouck dans l'exploitation d'une approche similaire portant sur d'autres contenus, comme nous l'avons relevé dans son portrait.

				mutualisable, mais demande encore à être transposé à d'autres contenus	
--	--	--	--	---	--

Le tableau suivant montre les mêmes données relativement à Alizée.

Tableau 19 : Indices de développement potentiel des compétences évaluatives d'Alizée

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Oui, référé aux contenus, reformulations à l'aide d'une taxonomie comme outil analytique	Oui, en rapport avec les objectifs dans une perspective analytique			
B	Oui, traduisent des objets clés de la discipline présents dans le PER	Oui, traduisent des objets clés de la discipline	Oui, le plus souvent complexes du point de vue des habiletés cognitives et des contenus (compréhension)		
C	Oui, les critères sont formulés en référence aux contenus issus du PER	Oui, les critères sont pensés depuis les contenus		Oui, approches critériées sur des apprentissages complexes	
D	Oui, la note tend à avoir un rapport avec les contenus issus du PER	Oui, la note rend compte d'éléments de contenus	Oui, la note traduit la complexité des objets présents dans les tâches	Oui, la note est issue d'un raisonnement référé aux unités d'apprentissag es pour le calcul du seuil du 4, et dans certaines situations, pour toutes les notes	Oui, la note n'est plus compensatoire, permet un feedback aux élèves
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures évitables
F	Oui, rapport distancé	Oui, choix assumés	Oui, choix assumés référés à une taxonomie	Oui, le système critérié est considéré comme mutualisable, mais demande encore à être	Oui, les directives en la matière sont critiquées, voire refusées

				transposé à d'autres contenus	
--	--	--	--	-------------------------------------	--

Nous l'avons montré dans notre partie conceptuelle sur l'AL, la perspective que nous en adoptons se traduit au travers de cinq caractéristiques, que nous reprenons pour montrer plus finement le développement des compétences évaluatives d'Anouck et d'Alizée. Pour la clarté du texte, nous les récrivons en italique dans la section suivante. Il nous semble intéressant de mettre ces différentes caractéristiques en dialogue avec nos résultats pour donner à voir le processus de développement des compétences des deux enseignantes.

5.3 Le processus de développement des compétences évaluatives chez chaque enseignante au regard des caractéristiques de l'assessment literacy

1. *La prise en compte du contexte dans lequel s'inscrivent leurs pratiques.* Les deux enseignantes ont collaboré durant tout le processus de recherche-formation. Leur contexte s'est notamment traduit au travers de cette collaboration, puisqu'elles ont sans cesse dû s'y référer pour élaborer leurs évaluations. À de nombreuses reprises, elles en ont questionné certaines dimensions - comme les normes imposées au niveau des barèmes, le nombre d'évaluations demandées par semestre, voire les contraintes liées au système informatique de gestion des notes imposé par le Département de la Formation du canton de Vaud – mais cela ne les a pas empêchées d'infléchir leurs pratiques. Alizée, par exemple, et à de nombreuses reprises, a questionné les directives relatives à la notation desquelles elle peut maintenant se distancer en pleine conscience.
2. *La négociation par Anouck et Alizée des objets disciplinaires en jeu dans l'enseignement, l'apprentissage et l'évaluation.* Elles choisissent les objets disciplinaires qu'elles vont traiter, et surtout la manière avec laquelle elles les enseignent, en fonction de plusieurs facteurs pensés de plus en plus en référence à l'alignement curriculaire élargi. Ce processus décisionnel, qui caractérise l'AL, nous fait penser à une action de négociation qu'Anouck et Alizée mènent à chaque fois qu'il s'agit de savoir ce que les élèves vont devoir apprendre, même si des questions subsistent au niveau de certains contenus.
3. *Leur prise en compte des processus d'enseignement et d'apprentissage pour penser les pratiques évaluatives.* Nous avons pu observer que chacune d'entre elle, dès le début, accorde une place importante aux processus d'enseignement et d'apprentissage dans ses

démarches évaluatives. Ce qui évolue selon nous, c'est ce sur quoi porte la réflexion. Alors qu'au départ elles semblent davantage être attentives à tout vouloir évaluer, elles opèrent progressivement des choix et acceptent de centrer leurs prises d'informations sur des contenus clés.

4. *Les connaissances incontournables en évaluation d'Anouck et Alizée référées à leur contexte.* Nous l'avons abondamment documenté, elles semblent s'être appropriées, à leur manière, la taxonomie comme un outil théorique leur permettant de mieux saisir le degré de complexité des contenus et des tâches et ainsi comprendre la cohérence de leurs pratiques évaluatives sommatives. Quant à l'alignement curriculaire élargi, nous faisons l'hypothèse qu'il leur a permis d'affiner leur connaissance du PER, des caractéristiques des tâches évaluatives, des systèmes de pondération et de notation, au vu de leurs propos mais surtout des épreuves créées en commun.
5. *Une pratique évaluative considérée comme relevant, pour Anouck et Alizée, d'un compromis permanent entre différentes dimensions.* Ces compromis sont très explicites lorsqu'Anouck exprime encore des doutes à expérimenter ses démarches qualitatives sur d'autres contenus, quand elle décrit le passage entre le travail de critérisation et celui de la notation comme restant délicat « *C'est compliqué de le faire, mais j'arrive mieux à me sentir légitime dans mon échelle ensuite* » (entretien 2, 25), ou encore lorsqu'elle reformule ou sélectionne des objectifs, tâche ô combien complexe. On le comprend également lorsqu'Alizée, sur le même sujet, dit : « *Je pense que là, pour moi, le plus complexe, c'était au niveau entre, la fin. Entre les points et l'échelle* » (entretien 2, 262).

5.4 En guise de synthèse provisoire

Au regard de ce qui précède, nous voyons que toutes les compétences se développent significativement.

Malgré certains doutes émis par Anouck et Alizée et un questionnement certes moins hésitant mais toujours ouvert, nous ne pouvons observer de tension sur une des catégories qui entraînerait de fait une résistance au développement de plusieurs de leurs compétences évaluatives. Ce constat, encore une fois, contraste avec ce qui a émergé des résultats des enseignants de mathématiques. Cela nous pousse à interroger le rôle que joue la discipline, voire certains objets disciplinaires, dans l'impact sur les pratiques évaluatives que peut provoquer l'appropriation du modèle théorique de l'alignement curriculaire élargi. Mais aussi à questionner l'impact du contexte sur la pérennité des changements dans les pratiques lorsque

ces dernières sont le fruit d'un travail collaboratif. Nous y reviendrons dans notre conclusion.

6 Portrait 6 : Samia, enseignante de français

La réalisation du portrait de Samia s'appuie sur les données suivantes :

- Le premier entretien et le premier écrit, réalisés tout deux avant la formation ;
- Sa première et seconde évaluation¹⁵⁷ ainsi que leur discussion proposée par le formateur, son deuxième écrit, tous produits durant la formation ;
- Les échanges entre Samia, ses pairs et le formateur durant les séminaires de formation 2 et 3.
- Le second entretien et le troisième écrit, réalisés à la fin de la formation.

Abordons maintenant les dimensions interrogées par nos questions et sous-questions de recherche.

6.1 Les dimensions des pratiques évaluatives de Samia sur lesquelles il porte son attention lorsque ces dernières sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi

Avant la formation

6.1.1 Les dimensions de la pratique évaluative sur lesquelles Samia porte son attention *avant la formation*

Le premier écrit de Samia rend compte d'un questionnement en situation. Elle explique combien être juste est important pour elle en évaluation. Elle exprime la volonté que ses élèves ne ressentent pas « *les sentiments d'injustice et d'humiliation* » qu'elle a rencontrés lorsqu'elle était à leur place. Pour cela, elle explicite, en prenant exemple sur l'évaluation de la production écrite, son emploi de grilles de critères pour être la plus objective possible. Mais lorsqu'elle corrige, elle constate que « *des textes relativement pauvres en imagination ou intérêt obtiennent quand même la moyenne grâce aux différents objectifs. Et c'est également parfois le contraire, des textes assez mauvais au niveau syntaxe et structuration obtiennent la moyenne grâce à la richesse des idées ou simplement au fait d'avoir suivi scrupuleusement les objectifs* », ce qui la

¹⁵⁷ Samia a réalisé la première évaluation seule. En revanche, elle a construit la seconde en collaboration avec sa collègue d'établissement Julie, participant également à la recherche, dont le portrait n'a pas été réalisé, faute de données représentatives (absence de participation dans les séminaires).

perturbe, puisque trop d'objectivité a ici tué l'originalité du texte. Aussi se demande-t-elle s'il ne faut pas réintroduire « *des points uniquement subjectifs* ». Une tension est donc perceptible entre objectivité et subjectivité, et, plus précisément, entre l'évaluation de la production écrite et le fonctionnement de la langue, source de doutes : « *je m'interroge beaucoup sur cette façon de faire finalement [...] des élèves qui sont bons en grammaire ou en orthographe, ils vont avoir quand même pas mal de point là-dessus, et s'ils arrivent pas bien à écrire, ou si la rédaction est pas terrible, ils ont quand même un 4, tout comme des élèves qui sont excellents en rédaction, mais qui ont beaucoup de peine en grammaire ; pis eux vont aussi quand même atteindre ce 4* ».

Cette recherche de justice, très présente dans son premier entretien « *« est-ce que je suis juste, est-ce que je suis injuste, est-ce que je suis sévère, trop sévère ? »* », est donc bien référée à cette évaluation de plusieurs objets comprise comme une norme : « *maintenant j'ai l'impression que dans les TS, tout est toujours mélangé, donc l'élève qui a de la peine, il peut pas se concentrer sur un objectif pis se dire « ah bah là j'ai réussi »*. On teste tout, plus ou moins constamment ». Là, Samia se demande si elle ne pourrait pas discerner les objets, mais rien n'est stabilisé.

Toutefois, ce questionnement important prend forme sur des pratiques dont certaines caractéristiques sont stabilisées :

- Tout ce qui est évalué a été travaillé en classe : « *Chacun de ces sujets est travaillé simultanément pendant la période de structuration, ceci de manière très scolaire, et pourra ainsi être évalué lors de la rédaction. En expression, les élèves rédigent énormément de petits textes, complètent des exercices liés à la structure du récit, à la manière de rédiger une scène de suspense, etc...* ». L'enseignement des objets de fonctionnement de la langue sont exploités en situation et mis en lien avec l'apprentissage complexe visé : « *dans ma manière de pratiquer, je vais me dire ok, je veux un TS de rédaction, qu'est-ce que je vais user comme grammaire là-dedans ? Alors là j'ai dit les groupes nominaux, l'accord sujet-verbe. Qu'est-ce que je veux en conjugaison ? Ah on est dans le passé, alors je mets le temps du récit au passé, la concordance des temps au passé. Qu'est-ce que je veux en vocabulaire ? Ah bah alors je vais entraîner là les synonymes, les répétitions, le registre de langue. Je mets ça dans ma grille, en plus du schéma narratif, etc. pour la rédaction. Ils la reçoivent au début, et puis on fait des exercices, on écrit, et chaque fois on relève...* » ;

- Le PER est le point de départ de ce raisonnement : « *en début d'année, j'ai analysé le PER, et choisi pour chaque sujet d'expression les sujets de structurations qui correspondaient le mieux* » ;
- La répartition des évaluations est fixe : « Donc je fais un TS par axe, plus ma cinquième note, c'est l'ensemble de mes petits tests [...] La cinquième note c'est le travail assimilé ; pis dans ce travail assimilé, j'ai une note de grammaire, une note de conjugaison, une note d'orthographe, une note de... ». Les TS sont « toujours complètement en lien expression-structuration » ;
- Les grilles d'évaluation, lorsqu'elles existent, sont données aux élèves avant la passation de l'épreuve ;
- Le moment d'élaboration de l'épreuve est souvent le même : « On fait jamais le test avant, en fait. A part la grille de rédaction. Oui, parce qu'on travaille avec. Les autres tests on les prépare pas avant. On attend toujours d'avoir bien avancé dans le sujet, et je dirais, aux deux tiers du temps, on fabrique le test. Parce que ça nous donne des idées aussi en travaillant avec les élèves, en étant en classe, on voit jusqu'où on a fait, et après on prépare le test ».

Plus spécifiquement, nous observons que Samia entretient avec le PER un rapport distancé. S'il est pour elle, comme nous l'avons déjà évoqué, une référence globale pour son enseignement « *c'est au début de l'année qu'il intervient ; maintenant je l'ouvre plus, parce que j'ai déjà repéré tous mes éléments* », il n'est pas une référence pour déterminer les objectifs à poursuivre : « *On utilise aussi beaucoup bêtement le livre unique et l'atelier du langage*¹⁵⁸ *pour tout ; les sujets sont dedans. Le PER on l'utilise principalement pour les quatre axes; quel est le but final, qu'est-ce qu'eux ils veulent qu'on ait atteint en fin de dixième en expression orale, qu'est-ce qu'ils veulent, voilà... Et pis après on avance...* ». En évaluation non plus, le découpage du PER reste pour Samia trop aléatoire, elle préfère se référer, encore, aux moyens ou à ses propres documents pour déterminer les objectifs plus précis en jeu : « *moi avec les élèves quand je leur annonce un test, on note plutôt les éléments... la page de théorie, on résume, on fait des résumés des théories et pis on dit « voilà c'est ça qu'il faut maîtriser ».*

Si un questionnement est à l'œuvre au niveau des objets de savoir et du plan d'études, le plus gros souci pour Samia se situe au niveau de la notation. D'entrée de jeu, en entretien, elle dit :

¹⁵⁸ Moyens d'enseignement de référence utilisés au moment de la recherche dans les degrés 9-11.

« je trouve que c'est un cadre qui est quand même pas très précis, par rapport aux échelles, par exemple ; c'est relativement flou ». Puis, elle étaye « Le barème c'est le gros problème justement [...] c'est un de mes gros soucis ». Là, elle rend compte de pratiques et de croyances qu'elle peut argumenter : « j'oscille entre 60 et 70% pour le 4. Pis suivant la difficulté ou la complexité de l'évaluation – pour moi la complexité ça veut dire qu'il faut être capable d'utiliser plusieurs notions simultanément, c'est ça qui fait la complexité – je descends le barème ; et plus c'est une évaluation juste ou faux basique, je le monte un peu. Mais de règle générale, je suis vers 66%, je me dis bon bah voilà, c'est deux tiers, ça me paraît logique que ça soit le 4, pis je le prends sur ibaremes.ch, comme ça il est régulier depuis en bas à en haut, je l'imprime et pis voilà ». Dans ce sens, un barème ne doit pour elle jamais être modifié. Elle fixe d'abord son seuil de suffisance, puis détermine, sans conviction, les autres notes : « Je fixe le 4, mon seuil, et après j'équilibre sur ibaremes ». En effet, ses pratiques de notation semblent être conditionnées par des prescriptions qu'elle donne l'impression de subir, sans pouvoir les interroger : « ce que la LEO... nous oblige de faire. Autrefois, on avait plus tendance à élargir en bas, pis à serrer en haut. Maintenant on peut plus, donc c'est complètement équilibré partout ». Noter de cette manière ne lui convient donc pas, la destabillise : « moi je me dis « est-ce que je vais chercher trop loin ? » ».

Cette dernière question rend compte selon nous d'une attention particulière donnée aux apprentissages des élèves, à ce qu'ils vont devoir maîtriser, et à comment valider leurs acquis. Les objets de savoir occupent donc dans son raisonnement une place privilégiée, même s'ils ne sont que peu considérés au travers du PER « vaut mieux tout réadapter ; le but, c'est qu'ils [les élèves] sachent ce qu'ils ont à apprendre ». Dans ce sens, une « bonne » évaluation est, pour elle, une épreuve sans consignes « C'est-à-dire, que l'élève s'est tellement approprié ce qu'il y avait à faire, ce qu'on lui demandait, qu'il démarre tout seul. Voilà. En gros. Mais c'est vrai, que bêtement, pour moi c'est ça. C'est qu'il a compris », qui est en rapport avec l'enseignement, et qui propose des exercices simples et complexes en vue de discriminer le 4 de l'excellence « je trouve qu'il faut avoir une certaine proportion de questions basiques, qui peuvent paraître simples, pour que ceux-là puissent les réussir, qu'on puisse vérifier que la base a été acquise. ».

Globalement donc, nous observons une enseignante passablement destabillisée par la prescription, qui rend compte par ailleurs de pratiques lui donnant satisfaction mais également sources de questions jugées fondamentales.

Pendant la formation

6.1.2 Les points de cohérence et de ruptures d'alignement observés par Samia dans ses pratiques d'évaluation sommative lorsqu'il se réfère au modèle de l'alignement curriculaire élargi *pendant la formation*

Rappelons ici brièvement les trois temps de la réponse à cette sous-question :

- Présentation des données brutes en nous centrant sur les parties qui ont suscité un débat ou qui mettent en évidence des zones de tensions, de discussions, et en respectant l'ordre chronologique dans lequel elles ont été produites (ici : épreuve 1, épreuve 2, écrit 2).
- Présentation d'un tableau de synthèse de ces éléments que nous interpréterons en référence aux catégories conceptualisantes que nous avons définies.
- Commentaires de nos résultats.

6.1.2.1 Première évaluation créée pendant la formation

Cette évaluation porte sur l'expression orale au travers du texte théâtral¹⁵⁹. Elle a été passée par des élèves de 10 VG niveau 1. Elle est constituée de deux parties. La première consiste en une approche « théorique » de « l'amour médecin », une œuvre de Molière. La deuxième évalue la production orale d'un extrait du texte de l'auteur (annexes Partie 4, 1.2.3). Reprenons-les dans l'ordre, en mettant chaque partie en correspondance avec les éléments de discussion proposés par le formateur¹⁶⁰ et discutés en séminaire.

La partie 1 de l'épreuve consiste en huit questions relatives à l'œuvre de Molière et plus largement au théâtre. À côté de chacune d'entre elles figure le nombre de points, 13 au total. Cette épreuve évalue l'objectif « L1-31: lire et analyser des textes de genres différents et en dégager les multiples sens ». Sur la page de garde du TS, Samia y a apparié quatre composantes et sept éléments de progression des apprentissages. Mais six objectifs, reformulés plus précisément par ses soins, sont également notés au bas de l'épreuve¹⁶¹.

Les éléments de discussion relevés par le formateur concernant cette partie sont les suivants :

¹⁵⁹ Samia explique toute sa démarche de création dans son deuxième récit dont nous parlons plus bas.

¹⁶⁰ Rappelons que nous sommes le formateur. Pour une discussion critique sur son rôle dans cette recherche, le lecteur se dirigera vers la conclusion finale de ce manuscrit.

¹⁶¹ Ce qui est cohérent avec les propos de Samia récoltés durant l'entretien.

- Un questionnement global sur le niveau de complexité des items : « vos items évaluent-ils de l'analyse ? » ;
- Une interrogation sur l'objet de savoir évalué : « Globalement, quelles caractéristiques principales du texte littéraire sont évaluées ici ? Sont-elles toutes présentes ? Évaluez-vous aussi autre chose ? ».

1) Donne 3 éléments bien distincts concernant la biographie de Molière:

a) _____

b) _____

c) _____

...../ 3pts

2) Deux grands écrivains ont vécu au même siècle que Molière: Jean de la Fontaine et Corneille. Quels genres de textes écrivaient-ils?

Jean de la Fontaine: _____

Corneille: _____

...../ 1pt

3) Une pièce de théâtre est divisée en _____ contenant eux-mêmes plusieurs _____.

...../ 1pt

4) Les indications de lieux, de mise en scène, de mouvements et de sentiments sont des _____. On les repère dans le texte parce qu'elles sont écrites en _____.

...../ 1pt

5) Dessine le schéma actanciel de *Lucinde* dans cette pièce:

...../ 2pts

Figure 51 : Samia, épreuve 1, expression orale au travers du texte théâtral, items 1 à 5

6) Les caractéristiques du genre

Cite 2 différences entre la comédie et la tragédie

COMEDIE	TRAGEDIE

..... / 2pts

7) Au théâtre, on parle de «double énonciation». Explique cette expression:

..... / 1pt

8) Site les 4 différents procédés comiques que l'on trouve dans les pièces de Molière

a) Le comique de.....

b) Le comique de.....

c) Le comique de.....

d) Le comique de.....

..... / 2pts



Figure 52 : Samia, épreuve 1, expression orale au travers du texte théâtral, items 6 à 8

Relativement à cette partie, d'autres éléments de discussion sont relevés sur la page de garde, ainsi que sur la grille de critères, à savoir :

- Une interrogation autour de ce choix d'évaluer des aspects théoriques : « Pourquoi évaluer cela ici sous forme théorique ? Ou alors, qu'est-ce qui vous empêche de mettre 2 notes ? » ;
- Un questionnement autour d'une possible double notation si ce choix perdurait (une pour L1-31 et l'autre pour L1-34);
- Un doute émis autour du nombre de progressions des apprentissages issues du PER possiblement évalués ici : « Évaluez-vous vraiment tout cela ? Vous en reprenez une partie dans vos objectifs, quid des autres ? » ;
- Une interrogation à inclure dans la grille les 13 points des aspects théoriques.

Relativement à la deuxième partie, et plus spécifiquement aux objectifs explicités, les éléments de discussion ont été les suivants :

- Une recherche de liens de cohérence entre la feuille de route et les progressions d'apprentissage du PER sélectionnées pour l'objectif L1-34 ;
- Une remarque relative à la pondération possiblement déséquilibrée entre les différentes progressions d'apprentissage du PER.

L1 34: (production)

- planification de la production orale en tenant compte des caractéristiques du genre théâtral et des préparations que cela suppose
- hiérarchisation des idées, préparation de la mise en scène, adaptation des supports (décors), adaptation de l'espace
- observation du protocole de la production orale propre au genre théâtral (écoute, entrées et sorties de scène, tour de paroles, salutations,...)
- adaptation du ton, du volume, de la diction, du rythme à la situation d'énonciation
- adaptation de la gestuelle
- expression des sentiments, émotions (L1 31 p.30)
- travail de mémorisation et restitution fidèle du texte appris

Figure 53 : Samia, épreuve 1, expression orale au travers du texte théâtral, progressions des apprentissages du PER sélectionnées pour l'objectif L1-34

Puis, concernant la feuille de route, le formateur interroge le statut des contenus qui y figurent.

Feuille de route

Ta feuille de route doit contenir les informations suivantes:

- 1) prénom
- 2) distribution des rôles
- 3) date de passage
- 4) délai choisi d'apprentissage du texte
- 5) organisation des répétitions (dates prévues /dates validées, présence des acteurs)
- 6) lieu des répétitions
- 7) bref descriptif du choix des costumes
- 8) liste des accessoires nécessaires
- 9) éléments clés de la mise en scène (entrée et sortie des acteurs, particularités prise en compte des didascalies...)
- 10) responsabilités de chaque membre du groupe, partage des tâches

Cette feuille de route est personnelle et doit être **faite et rendue individuellement** au plus tard le jour de passage.
Sa présentation est libre, mais doit être claire et complète.

Figure 54 : Samia, épreuve 1, expression orale au travers du texte théâtral, feuille de route pour l'évaluation de l'objectif L1-34

Enfin, relativement à la grille d'évaluation, le formateur questionne la pondération (comment mettre 0, 1, 2 ou 3 ?), le regroupement de certains critères et la difficulté d'en évaluer d'autres (entrée et maintien dans le rôle, charge émotionnelle du comédien ; investissement personnel, dépassement de soi), et demande, avec un clin d'œil, pourquoi il n'y a pas d'échelle...

Grille d'évaluation

Critères	Points
la feuille de route est rendue dans les délais	0 1
la feuille de route est complète, claire et bien présentée	0 1 2 3
les responsabilités sont équilibrées, chaque membre a participé, collaboration	0 1 2
apprentissage par coeur du texte	0 1 2 3
interaction avec les autres personnages, écoute, regards,rythme	0 1 2 3
prise en compte du public (salutations, ne pas tourner le dos,apartes,...)	0 1 2
costumes et accessoires	0 1 2
gestuelle	0 1 2 3
ton	0 1 2 3
volume de la voix	0 1 2
mise en scène (entrées, sorties, tours de parole, occupation de l'espace)	0 1 2 3
entrée et maintient dans le rôle, charge émotionnelle du comédien	0 1 2 3
diction	0 1 2
respect des consignes de l'auteur, de l'époque et du genre de la comédie	0 1 2
investissement personnel, dépassement de soi	0 1 2
créativité, originalité	0 1
questionnaire théorique	 / 13 pts
Total:	 / 50 pts

Figure 55 : Samia, épreuve 1, expression orale au travers du texte théâtral, grille d'évaluation

La discussion est donc volontairement amenée vers toutes les composantes de l'alignement curriculaire élargi. Il est par ailleurs intéressant de noter que, lors du séminaire 2, Samia explique les raisons de son choix de vouloir évaluer théorie et production séparément : « *dans le PER, quand on va voir sous théâtre, ça regroupe beaucoup de théorie dans L1-31, des éléments dont ils ont besoin pour jouer une scène, raison pour laquelle je les introduis dans la grille, pour pas casser l'alignement* ». Mais elle dit « *avoir des doutes* ».

6.1.2.2 Seconde évaluation créée pendant la formation

La seconde évaluation est particulière¹⁶² (annexe Partie 4, 1.2.4). En effet, Samia décide d'aborder un contenu relatif au fonctionnement de la langue, qui prendra forme dans un TA¹⁶³. Trois semaines avant le dernier séminaire, elle écrit le message suivant au formateur : « *Nous avons décidé de travailler sur un ETA¹⁶⁴ de conjugaison, qui, jusqu'à ce jour, nous paraissait le plus simple à préparer. Or en réfléchissant sur le problème de compensation, nous avons tenté de décortiquer chaque exercice afin que l'élève sache exactement ce qu'il a compris ou pas. Cela revient à un travail de composition qui nous a pris plus de 5 heures, pour un simple ETA. La correction semble également plus difficile au final. Cependant, il est certain que de cette manière, chaque objectif est clairement vérifié et pondéré. Reste cependant le problème final, pour mettre la note, la compensation finit bien par intervenir... Nous pensons que cela reflète aussi la vie de tous les jours. Finalement, nous sommes tous plus doués pour une chose ou pour une autre, et l'important est de trouver un équilibre.... tout en sachant où se trouvent nos forces et nos faiblesses... Tout comme dans ce travail.....* » (correspondance personnelle, 13 avril 2015). Nous y reviendrons plus loin.

Cette épreuve porte donc sur de la conjugaison, plus spécifiquement sur l'évaluation du subjonctif présent et passé ainsi que sur l'impératif présent et passé. Elle se destine de nouveau à des élèves de 10G1, comme la précédente. Elle est constituée d'une page sur laquelle figurent l'objectif du PER « L1-36 : analyser le fonctionnement de la langue et élaborer des critères d'appréciation pour comprendre et produire des textes », six de ses progressions des apprentissages choisies par Samia et mises en rapport avec les exercices, ainsi que la liste des

¹⁶² Pour rappel, Samia l'a réalisée avec Julie, sa collègue d'établissement.

¹⁶³ Avec ce choix, Samia a dérogé à la règle implicite qui voulait que les enseignantes produisent des TS (Travaux significatifs) en référence à des tâches complexes. Mais pour rester cohérent avec le principe de recherche collaborative, et pour ne pas interférer avec ses questions, nous avons respecté ce choix, pensant finalement qu'il apporterait son lot de réflexions pertinentes.

¹⁶⁴ ETA : Élément de Travail Assimilé, terme usité dans l'établissement, synonyme de TA (Travail assimilé).

verbes travaillés en 10e et évalués. Notons également que des objectifs reformulés par elle et Julie sont également présents, afin d'aider les élèves à appréhender le contenu en jeu dans l'épreuve.

Progressions d'apprentissage:

L1 36:

- identification du système et de la valeur du subjonctif présent et passé (ex.5)
- identification du système et de la valeur de l'impératif présent et passé (ex.5)
- identifie les verbes 10H aux 4 temps travaillés (ex.1)
- construit et orthographie correctement les verbes de 10H aux 4 temps travaillés (ex.2)
- identifie l'impératif présent et le subjonctif présent en contexte (ex.3)
- construit et orthographie les verbes les plus fréquents en contexte aux 4 temps travaillés (ex.4)

Figure 56 : Samia, épreuve 2, conjugaison, progressions des apprentissages du PER sélectionnées pour l'objectif L1-36

Un élément de discussion proposé par le formateur interroge ici le poids donné à chaque progression. Il questionne le déséquilibre présent.

Le contenu de l'épreuve, quant à lui, consiste en cinq exercices mis en lien chacun avec un des objectifs reformulés. Pour chacun d'entre eux, Samia a précisé la manière d'attribuer les points ainsi que leur total. L'ensemble de l'évaluation comprend 43 points. À la fin, le total par exercice est précisé et mis en correspondance avec une rubrique « objectifs atteints » qui permet de calculer le seuil de réussite du 4, à savoir 27 points. Nous reproduisons chaque exercice en intégralité vu qu'ils ont chacun fait l'objet d'une discussion.

Exercice 1:

Indique, dans le tableau suivant, si le verbe conjugué est au :

subjonctif présent / subjonctif passé / impératif présent / impératif passé / autre

verbe	temps
que je serve	
envoie	
que tu eusses mangé	
dormez	
que j'aie dormi	
que nous partions	
ayez fini	
que j'aie couru	
vous pesiez	
sois assis	
tu manges	
que nous jetassions	

Objectif 1 : identifier les verbes aux 4 temps:

subjonctif présent : 0 - 1 - 2

subjonctif passé : 0 - 1 - 2

impératif présent: 0 - 1 - 2

impératif passé : 0 - 1 - 2

Total exercice 1 : / 8pts

Figure 57 : Samia, épreuve 2, conjugaison, exercice 1

Là, le formateur demande d'expliciter la genèse du système de pondération.

Exercice 2: Complète les tableaux suivants:

	subjonctif présent	impératif présent
recevoir, il		
manger, tu		
rire, nous		
envoyer, vous		
agir, je		
céder, tu		
élever, nous		
servir, tu		
cuire, ils		
plaire, nous		
Total par temps	 / 10 =pts	 / 10 =pts

	subjonctif passé	impératif passé
s'asseoir, nous		
payer, tu		
craindre, vous		
être, je		
avoir, nous		
espérer, vous		
partir, tu		
dormir, nous		
s'écrier, tu		
oublier, elle		
Total par temps	 / 10 =pts	 / 10 =pts

Objectif 2: construis et orthographe correctement les verbes de 10H aux 4 temps
travaillés

0-2pts = 0pt

3-5pts = 1pt

6-8pts = 2pts

9-10pts = 3pts

Total exercice 2: / 12pts

Figure 58 : Samia, épreuve 2, conjugaison, exercice 2

Ici, le formateur encourage à utiliser le terme de « juste » pour attribuer les points.

Exercice 3: dans le texte suivant, souligne

en : vert les verbes au subjonctif présent

en: rouge les verbes à l'impératif présent

La météo annonce de la neige pour les jours prochains. Il se peut que j'aille passer quelques jours dans une station, alors, il faut absolument que je sorte mes vêtements de ski, que je les trie et essaie ma combinaison. Je descends chercher mes affaires à la cave.

- Dépêche-toi!, me prévient ma maman
- Je ne trouve rien dans cette cave!
- Cherche mieux et ouvre les yeux! Nous avons tout rangé l'hiver passé.

Cherche et trouve,...me dirait mon professeur de français... Après quelques minutes, je retrouve enfin mes affaires.

Je remarque qu'il me manque un casque. Il faudra que je m'en achète un! Quel travail!

Qu'il fasse beau ou qu'il pleuve, cela m'est souvent égal. Excepté pour le ski!

Mais aujourd'hui, qu'il fait beau!!!! Je pense qu'il fera le même temps à la montagne. Du moins je l'espère!

En arrivant sur place, il faudra aussi que je loue des skis, que je souscrive une assurance et que je paie un abonnement. Bref, que je n'oublie rien! Pensez à prendre votre tête, me répétaient souvent mes grands-parents. Oh! Combien ils avaient raison! Revenez m'aider, s'il vous plaît! Je vais tout oublier.....Soyez gentils...

Objectif 3: identifier l'impératif présent et le subjonctif présent dans un texte

subjonctif présent : justes : / 12 faux : pts:

impératif présent : justes: / 8 faux : pts:

Total : justes - faux divisés par 2 (pas de demi point)

Total : / 10pts

Figure 59 : Samia, épreuve 2, conjugaison, exercice 3

Dans l'exercice 3, c'est le calcul du total de points qui est questionné.

Exercice 4 : complète les phrases suivantes à l'aide du temps qui convient
(subjonctif présent + passé / impératif présent + passé)

- 1) Il faut que tu (être) _____ à l'heure à ton cours.
- 2) Les enfants, (rire) _____ moins fort. Bien que votre petit frère
(dormir) _____, il pourrait vous entendre.
- 3) Marie souhaite qu'ils (partir) _____ bientôt, il est temps!
- 4) Ne (jouer) _____ pas sous la pluie, sinon tu seras complètement mouillée.
- 5) Pierre et Paul, (manger) _____ vos glaces immédiatement.
- 6) J'irai à la fête pourvu que Jeanne ne (venir) _____ pas.
- 7) Michel est furieux que son frère (oublier) _____ son anniversaire.
- 8) Ne (courir) _____ pas sur la pelouse, nous pourrions écraser les fleurs.
- 9) Bien que les filles (oublier) _____ l'heure, elle sont arrivées à temps.
- 10) Je doute que tu (pouvoir) _____ finir ton travail avant ce soir.

Objectif 4: construit et orthographie les verbes les plus fréquents en contexte

choix du temps : 0.5pts / conjugaison : 0,5pts

Total: / 11pts

Figure 60 : Samia, épreuve 2, conjugaison, exercice 4

Dans cet exercice 4, un questionnaire intervient autour du rapport entre les éléments de progression des apprentissages du PER et les dimensions sur lesquelles portent les points. Puis, dans l'exercice 5, c'est sa faible pondération qui est remise en question.

Exercice 5: complète

Le mode du _____ est utilisé pour exprimer l'incertitude ou la possibilité d'une action. L'impératif permet d'exprimer _____ , _____ ou _____ .

Total: / 2pts

Figure 61 : Samia, épreuve 2, conjugaison, exercice 5

Enfin, relativement à l'échelle, le formateur questionne le phénomène de compensation possible dans le calcul de la note, et cela même si une recherche de seuil par exercice lui semble pertinent.

Pondération:

Objectifs atteints (4)

Exercice 1: 5 pts

Exercice 2: 8 pts

Exercice 3: 5.5 pts

Exercice 4: 6.5 pts

Exercice 5 : 2 pts

Total : 27 / 43pts

Echelle:

1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
0-6.5	7-11.5	12-16.5	17-21.5	22-26.5	27-30.5	31-33.5	34-37.5	38-40.5	41-43

Figure 62 : Samia, épreuve 2, conjugaison, échelle

En début de séminaire 3, Samia explique à ses collègues que le but de cet exercice était pour

elle de voir dans quelle mesure il était possible d'éviter le phénomène de compensation en évaluant le fonctionnement de la langue. Elle regrette ne pas y être arrivée, ce qui donnera lieu à des débats dont nous rendons compte plus bas.

6.1.2.3 Deuxième écrit réalisé pendant la formation

Son texte se réfère à l'évaluation discutée lors du séminaire 2 (annexe Partie 5, 7.8). Elle y détaille les réflexions, questions, voire dilemmes qu'a suscité son élaboration.

Premier élément important, une prise de conscience semble se réaliser par rapport aux pratiques habituelles : *« Pour évaluer cela [la production orale au travers du théâtre] je faisais une simple grille, dans laquelle je tenais compte de l'apprentissage par coeur, du ton et du rythme de la voix, de l'expression gestuelle, et des costumes. Après notre séance de formation, en regardant dans le PER, je me suis aperçue que cela ne correspondait pas réellement à des composantes du PER. C'est seulement en fouillant, que j'ai réalisé que beaucoup de composantes et de progressions d'apprentissages étaient travaillées mais pas évaluées ! Je n'évaluais que la production finale et non pas les tâches relatives à cette dernière. Il y avait déjà une cassure dans l'alignement curriculaire ! »*. Les objectifs du plan d'études sont donc davantage pris en compte, les conditions de réalisation également.

Samia reste donc attentive aux contenus. Elle se demande toujours et à plusieurs reprises *« qu'est-ce que j'évalue ? »* traduisant ainsi ses doutes sur le mélange des objets dans son épreuve (théorie et situation de communication) : *« Mais au final, est-ce réellement une évaluation d'expression orale ? Il y a aussi de la compréhension de l'écrit...! Est-ce possible de séparer les deux...? »*.

Cela étant, le fait de construire un seuil de suffisance en fonction des attentes relatives aux contenus, et de fonctionner avec une grille de critères permettant de rendre visible les exigences est pour elle une avancée significative : *« grâce à la grille et cette feuille de route, j'ai pu mettre mes résultats bien plus facilement que d'habitude, et surtout plus sereinement... . Le barème choisi convient parfaitement »*. Et de conclure : *« ce procédé est positif. Il m'a aidée à me sentir plus «juste» dans mon évaluation, et a permis, à certains élèves, de savoir où ils allaient, et comment s'y prendre »*.

6.1.3 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives de Samia à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *pendant la formation*

Sont prises en compte ici trois des quatre sources de données particulièrement significatives durant la formation :

- Les éléments de discussion proposés par le formateur (annexe Partie 5, 9). Mentionnés plus haut, ils se réfèrent tous au modèle théorique de l'alignement curriculaire élargi. Ils sont reproduits dans le tableau ;
- Les éléments débattus en séminaires de formation 2 et 3, retranscrits depuis les enregistrements ;
- Le deuxième écrit documenté précédemment.

Le tableau ci-dessous rend compte de la triangulation de ces données référée à nos catégories conceptualisantes.

Tableau 20 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Samia durant la formation

Données Catégories	Evaluation 1 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Samia débattus en séminaire 2 de formation	Evaluation 2 : éléments de discussion proposés par le formateur	Eléments acquis et/ou de discussion pour Samia débattus en séminaire 3 de formation	Ecrit 2
Rapport au référentiel	Toutes les progressions des apprentissages sont-elles évaluées ?	« Dans le PER, il n'y rien » « je ne vais pas rouvrir le PER de si tôt, j'en suis dégoûtée »	Reformulation des objectifs du PER à l'aide de la taxonomie ?		Nouvelle exploitation du PER pour aller chercher les contenus correspondant au sujet
Expression de l'objet	Quel contenu est évalué ici ?	N'y a-t-il pas trop de contenus évalués ?			Réflexion sur « qu'est-ce que j'évalue ? »
	Pourquoi évaluer une partie théorique ? Les items évaluent-ils de l'analyse ?	Quel statut donner aux objectifs transversaux ?			Possible de séparer expression orale et compréhension de l'écrit ?

Elaboration ¹⁶⁵ des tâches évaluatives	Cohérence entre la feuille de route et les progressions d'apprentissage du PER sélectionnées pour l'objectif L1-34 ?				
Pratiques de pondération	Pertinence d'inclure la théorie comme critère ?	Une réflexion sur le seuil du 4 en référence aux contenus est réalisée (les points en gras sur chaque critère)	Questionnement du poids donné à chaque progression des apprentissages du PER		Attribution des points correspondant aux progressions d'apprentissage du PER
	Pondération possiblement déséquilibrée entre les différentes progressions d'apprentissage du PER	Revenir vers des critères spécifiques en référence aux contenus			
	Comment mettre 0, 1, 2 ou 3 ?	Réflexion autour d'un minimum exprimé en référence aux progressions d'apprentissage du PER	Système d'attribution des points pas toujours transparent		Attribution des points minimiaux en fonction des contenus
	Regroupement de certains critères et la difficulté d'en évaluer d'autres				
Pratiques de notation	Qu'est-ce qui vous empêche de mettre 2 notes ?	Deux objectifs = deux notes : à discuter encore	Compensation toujours possible	À quelles conditions la compensation peut-elle être acceptable pour construire la note ? Comment gérer une construction de notes référées à l'apprentissage ?	Elaboration des seuils (le 4) plus juste, barème adéquat
	Pourquoi pas d'échelle ?	L'impact du contenu sur la possibilité de construire une échelle référée aux critères			
		Les pratiques pour éviter le phénomène de compensation : différent selon les contenus			

¹⁶⁵ Nous utilisons le terme de *design* de tâches dans notre modèle théorique, fidèlement à la littérature consultée, mais il nous semble plus parlant ici de parler d'élaboration, les deux étant synonymes.

Cette analyse au travers des catégories conceptualisantes donne à voir sur quelles dimensions de sa pratique Samia, avec l'aide de ses pairs et du formateur, porte ses réflexions. Malgré le fait que sa situation soit unique (elle est la seule à avoir élaboré individuellement un premier TS, puis en équipe un TA, sur des objets par ailleurs fort différents), ses mouvements de pratiques nous semblent révélateurs de réflexions intéressantes.

Par exemple, nous observons que le PER est toujours considéré comme peu utile en évaluation, mais qu'elle l'exploite désormais à des fins de reformulation des objectifs pour que ces derniers soient mis en rapport avec les exercices des épreuves.

Cette nouvelle réflexion n'est peut-être pas sans rapport avec une attention grandissante sur les contenus à enseigner, et surtout à évaluer, dans une recherche permanente de « ce qui doit être maîtrisé ». À ce niveau, Samia continue à investiguer son questionnement de départ autour de la multiplication des objets dans l'évaluation, et donne l'impression d'y voir maintenant plus clair.

C'est hypothétiquement la raison pour laquelle elle porte également une attention aux conditions permettant aux élèves de percevoir ses attentes et exigences (par exemple au travers de la feuille de route présentée en séminaire 2), ou ce qui lui fait désormais penser l'attribution des points et le calcul du seuil du 4 en référence aux contenus. Là, elle est parfois en mesure de faire des liens entre la pondération choisie et les objectifs du PER, ce qui est nouveau pour elle.

Mais un objet de débat, intense, perdure : celui de la compensation. Est-elle acceptable ? Si oui, à quelles conditions ? Peut-on admettre qu'un élève ne réussisse pas parce qu'il a montré des lacunes dans des aspects clés de l'apprentissage alors qu'il a fait preuve d'acquis ailleurs ? Comment faire face à ces dilemmes ? Là, Samia ne peut trancher. Et même si Anouk lui dira que c'est à elle de répondre, à elle de faire ses choix, Samia, bien que rassurée sur bien des aspects de ses pratiques, dira qu'elle ne peut assumer.

À la lumière de ces éléments, nous sommes en mesure de formuler quelques hypothèses interprétatives traduisant des mouvements dans les pratiques évaluatives de Samia, et qui nous permettent de voir dans quelle mesure elle parvient à travailler leur cohérence du point de vue du modèle théorique :

- Le PER est une référence au niveau des progressions des apprentissages qui méritent toutefois d'être reformulées ;

- Les tâches sont référées à des contenus et des apprentissages ;
- La pondération est réalisée en fonction des contenus ;
- La notation, même issue d'un total de points, est davantage référée à l'apprentissage, surtout pour le calcul du seuil de suffisance (4).

Conjointement, il se dégage également une forte tension ou difficulté à dépasser des ruptures d'alignement curriculaire élargi :

- Le problème à accepter la compensation comme interférant la cohérence de l'évaluation, et l'impossibilité d'entrevoir des pratiques permettant de passer de critères à une note sur une échelle à onze positions.

Après la formation

6.1.4 Hypothèses interprétatives des mouvements opérés dans les pratiques évaluatives d'Samia à des fins d'identification des points de rupture et de cohérence d'alignement curriculaire élargi *à la fin de la formation*

Samia a produit son troisième écrit huit jours après le dernier séminaire de formation. Son récit, succinct, est rédigé sous forme de bilan. Il commence par une prise de conscience importante sur sa posture d'évaluatrice : *« Dès le début de la formation, la théorie me paraissait très claire et la réflexion logique. Il m'a fallu cependant du temps pour que mon «inconscient» se l'approprie. Effectivement je me suis aperçue combien les fausses croyances étaient ancrées en moi. J'ai dû me détacher de celles-ci, accepter de tout remettre en question. Et comme tout chamboulement; cela fait peur. Je ne savais plus à quoi me rattacher. La plus grande question qui s'est posée à moi : suis-je en accord avec ma manière de fonctionner ? Et la réponse a été : non ! J'étais un brave petit «mouton» qui suivait les directives, sans trop réfléchir. Je ne me donnais à aucun moment le droit de me faire confiance, et de me regarder en tant que professionnelle. Je pense que pour moi le plus grand changement est là. Et gagner en liberté, c'est magnifique, mais aussi vertigineux ! »*. Une nouvelle assise semble donc être possible, responsabilisante, rassurante. Elle s'appuie sur les changements suivants :

- L'évaluation doit respecter l'enseignement et *« la perception de la réussite »* de Samia ;
- L'évaluation se centre sur ce qui est pertinent et constructif ;
- La réussite est envisagée en référence aux contenus : *« je peux maintenant m'autoriser à définir ce que j'estime être un minimum »* ;

- Une attention particulière est portée sur les progrès des élèves, sur leurs capacités à en être les acteurs.

Le tableau ci-dessous rend compte des éléments significatifs de deux sources de données relativement aux points de cohérence et/ou de rupture d'alignement curriculaire élargi à débattre : le dernier écrit et le second entretien. Ils sont directement mis en rapport avec les catégories conceptualisantes, de manière à rester synthétiques et centrés sur nos questions de recherche. Certes, des redondances peuvent apparaître, mais elles doivent à nos yeux être comprises et interprétées comme des signes d'un processus complexe d'appropriation de divers éléments par Samia. Nous commentons ces données plus bas en les mettant en dialogue avec les autres données.

Tableau 21 : Eléments disponibles pour l'analyse des points de cohérence et de rupture d'alignement curriculaire dans les pratiques évaluatives de Samia à la fin de la formation

<i>Données</i>	<i>Ecrit 3</i>	<i>Entretien 2</i>
Catégories		
Rapport au référentiel		Le PER est une ligne directrice, mais « <i>il est beaucoup trop large</i> »
		Je rajoute des objectif si le PER est incomplet : « <i>j'ai une capacité à le compléter</i> »
		Le PER incomplet pousse à réfléchir les contenus et leurs liens
		« <i>une aide à formuler les objectifs en structuration, mais en expression... ça je connais les... les grandes lignes. Donc là, je l'utilise pas beaucoup</i> »
		« <i>Je sais comment naviguer [dans le PER]</i> »
		Un flou persiste entre compétences et progressions des apprentissages
		Pas besoin de tout évaluer ; « <i>j'évalue vraiment les objectifs annoncés</i> »
Expression de l'objet	Ne pas tout évaluer, faire des choix, ne garder que ce qui est pertinent et constructif	« <i>il faut choisir les contenus qu'on évalue</i> »
	« <i>Que fait-on de la progression individuelle ?</i> »	« <i>Moins d'objectifs, et plus séparés les uns des autres, quand même. Pour pouvoir voir vraiment qu'est-ce qui est compris et qu'est-ce qui n'est pas compris</i> »
	Faire en sorte que l'élève construise son apprentissage	Nouvelle manière d'évaluer la technique en production écrite
		Plus facile d'évaluer car centration sur les contenus
		Forte attention au rapport contenu-tâche
		Analyse de l'objet plus récurrente

		Prise de conscience que comprendre n'est pas analyser
Elaboration des tâches évaluatives	« construire les exercices des évaluations de manières plus cohérente avec mon enseignement »	« une autre manière de considérer les TA »
		Le PER : « une aide à la fabrication des exercices »
		Des exercices plus ciblés, mais « pas évident pour les compréhensions écrites »
		« On ne peut pas avec un tout petit objectif, dire « l'élève a compris ». Donc vaut mieux prendre moins d'objectifs, faire moins d'exercices... »
		Moins d'exercices
Pratiques de pondération	« Je définis ce que j'estime être un minimum »	« Je vais faire en sorte que les points valident que l'objectif que j'ai choisi. De manière à vraiment pouvoir voir qu'est-ce qui est compris, qu'est-ce qui est pas compris »
		Les exercices basiques avec plus de points que les complexes
		« Les grilles, c'est plus facile parce que... soit j'évalue pas, soit je mets peu de points »
		Réflexion en termes de minimaux pour avoir les points qui font le 4
		« une pondération propre à la progression »
Pratiques de notation		Le raisonnement pour aboutir à ce soixante-six pourcents (seuil du 4) est complètement différent
		Notation en fonction de ce qui est maîtrisé
		« j'essaie d'avoir quand même, à partir de ce 4, une équité en haut et puis une équité en bas. C'est pas la même, hein, en haut et en bas »
		Le barème en amont permet de repenser les exercices

Ce tableau nous permet de formuler des hypothèses interprétatives quant aux compétences qui, progressivement, semblent se développer au travers d'une recherche de cohérence d'alignement curriculaire élargi¹⁶⁶ :

- Le PER est une balise à long terme, dont le rôle principal est d'être reformulé pour donner des repères en termes d'enseignement, d'apprentissage et d'évaluation. À ce titre, il est une référence pour re-construire du contenu ;
- Du point de vue des objets de savoir, trois éléments interpellent : ils sont considérés comme devant être peu nombreux dans l'évaluation afin d'être mieux jugés, ils sont à la base du raisonnement visant à circonscrire ce qui est important à maîtriser, et leur rapport aux tâches est souvent cité comme primordial ;

¹⁶⁶ Notre définition de la compétence et son statut dans la recherche sont explicités dans la partie conceptuelle sur l'*assessment literacy*. Nous l'intégrons progressivement dans nos résultats au regard de nos questions de recherche.

- Ces réflexions ont un lien direct avec les mouvements entrepris au niveau des tâches : elles sont moins nombreuses, référées aux contenus et objectifs, et conditionnent la pondération ;
- Les points, qu'ils soient indépendants ou reliés à des critères, sont pensés également en référence aux objectifs ;
- Le seuil du 4 est finalement construit au regard de ce qui est minimalement attendu dans les épreuves.

S'en dégagent également des tensions, ou difficultés à dépasser des ruptures d'alignement curriculaire élargi :

- Au niveau du PER, où la reformulation des objectifs et leurs ajouts demandent encore à être justifiés, transposés à d'autres objets, voire référés systématiquement à une taxonomie ;
- Relativement à la notation, où une attention à une forme de linéarité des échelles est encore présente.

Malgré ces derniers constats, nous faisons l'hypothèse que l'alignement curriculaire élargi apparaît ici comme une lunette analytique compréhensive aidante. Samia, dans son second entretien, y fait d'ailleurs référence : *« Alors au début, ça m'a paru compliqué au niveau de la terminologie. Bien que la plupart des mots, je les ai eu déjà entendus, je dirais, mais... Il a pris du sens quand je l'ai utilisé pour faire ma première évaluation, en fait. Quand on a dû faire l'exercice « créer une évaluation » en faisant attention, en le suivant, moi je l'ai vraiment mis, mais je l'avais devant moi, quoi. Je l'avais sur mon bureau et je prenais point par point. Et là, les... quand vous expliquiez ce qu'étaient les cassures, je... je comprenais, mais ça n'avait pas forcément de sens pratique... Et puis ben en fabricant mon évaluation, ben là, j'ai pu voir les cassures, en fait. Et là, c'est là que je l'ai vraiment intégré et compris, cet alignement »* (entretien 2, 147-148). Des propos qui se retrouvent dans ses deux évaluations, où une attention est portée à une recherche de cohérence entre chaque composante du modèle.

Tout au long de la recherche

En référence à ce que nous venons de mettre en évidence, nous faisons l'hypothèse que Samia est davantage outillé pour éviter des désalignements entre les objectifs, les tâches et la pondération. La notation a vraisemblablement gagné en cohérence, mais elle reste, avec la question de la compensation, le lieu de tensions non résolues. Il lui reste maintenant à en

explorer d'autres facettes, relativement à d'autres situations et contenus. Mais Samia a gagné en confiance.

Nous pouvons dès lors appuyer nos hypothèses interprétatives par les divers mouvements importants en œuvre dans les pratiques évaluatives sommatives de Samia décrits plus haut, que nous considérons comme autant d'évolutions rendues possibles par certaines composantes du modèle de l'alignement curriculaire élargi.

Mais qu'en est-il maintenant du point de vue du développement potentiel des compétences évaluatives de Samia, au regard de ces mouvements de pratique contrastés ?

6.2 Les compétences évaluatives que l'alignement curriculaire élargi participe à développer chez Samia

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Samia au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe. Pour répondre à cette question, nous mettrons en correspondance les apports déjà dégagés à l'aide des catégories conceptualisantes avec les compétences construites pour notre étude dont le développement est présenté dans la partie méthodologique.

6.2.1 Les compétences évaluatives développées principalement par Samia

Pour rappel, en lien avec notre deuxième question générale de recherche, nous visons ici à problématiser le développement des compétences en *assessment literacy* (AL) qui hypothétiquement se développent chez Samia au travers d'un travail interactif et réflexif sur le modèle théorique de l'alignement curriculaire élargi en alternance avec des expériences en classe. Pour répondre à cette question, nous mettrons en correspondance les apports déjà dégagés à l'aide des catégories conceptualisantes avec les compétences construites pour notre étude dont le développement est présenté dans la partie méthodologique.

Il nous semble ici important de reprendre chacune des six compétences et d'en justifier le possible développement. Pour éviter de charger le texte, nous rappelons qu'elles sont décrites dans le portrait de Fernand qui ouvre ce chapitre. Aussi, en nous appuyant sur nos résultats précédents, nous pouvons documenter quelles compétences Samia semble avoir développé :

Tableau 22 : Indices de développement potentiel des compétences évaluatives de Samia

	<i>Rapport au référentiel</i>	<i>Expression de l'objet</i>	<i>Elaboration des tâches évaluatives</i>	<i>Pratiques de pondération</i>	<i>Pratiques de notation</i>
Compétences					
A	Plus ou moins, la reformulation des objectifs devant être stabilisée	Oui, l'analyse de l'objet se fait de plus en plus en rapport avec le PER			
B	Oui, traduisent des contenus jugés importants	Oui, traduisent des objets clés de la discipline	Plus ou moins, leur niveau de complexité pouvant varier dans une même épreuve et leur rapport au PER pouvant être notoire		
C	Oui, les points sont donnés en référence aux contenus issus du PER			Plus ou moins, les critères étant systématiquement convertis en points et la compensation omniprésente	
D	Oui, la note 4 tend à avoir un rapport avec les contenus issus du PER	Oui, la note 4 tend à rendre compte d'éléments de contenus	Oui, la note 4 traduit la complexité des objets présents dans les tâches	Oui, la note est issue d'un raisonnement référé aux unités d'apprentissages pour le calcul du seuil du 4	Non, la note est toujours globalement compensatoire, ruptures probables
E	Ruptures évitables	Ruptures évitables	Ruptures évitables	Ruptures probables	Ruptures pour les notes au-dessus et au-dessous du 4
F	Oui, pour autant que la réinterprétation du PER soit proche de la source	Oui, choix assumés	Oui, choix assumés	Non, la compensation induite par les points semblant incontournable	Oui pour le 4, problématique pour les autres notes

Nous l'avons montré dans notre partie conceptuelle sur l'AL, la perspective que nous en adoptons se traduit au travers de cinq caractéristiques, que nous reprenons pour montrer plus finement le développement des compétences évaluatives de Samia. Pour la clarté du texte, nous les récrivons en italique dans la section suivante. Il nous semble intéressant de mettre ces différentes caractéristiques en dialogue avec nos résultats pour donner à voir le processus de développement des compétences de Samia.

6.2.2 Le processus de développement des compétences évaluatives chez Samia au regard des caractéristiques de l'assessment literacy

1. *La prise en compte du contexte dans lequel s'inscrivent ses pratiques.* Au départ, Samia identifie certaines directives comme autant de contraintes qu'elle ne comprend pas toujours. Elle met sur le compte de la LEO l'obligation de suivre certaines pratiques, comme le nombre de TS, leurs contenus, ou l'exploitation de barèmes standardisés. Elle déploie ses réflexions en en tenant compte, puisque l'on observe que des possibles sont envisageables, mais que des tensions demeurent, notamment au niveau des échelles pour qualifier les bons résultats.
2. *La négociation par Samia des objets disciplinaires en jeu dans l'enseignement, l'apprentissage et l'évaluation.* Le fait que Samia prenne des initiatives sur son interprétation du PER, allant même jusqu'à en récrire certaines parties lorsqu'elle en ressent le besoin, doit-il être compris comme un signe de développement professionnel ou plutôt comme un désarroi ? Nous pencherons pour la première hypothèse, partant de l'idée que ses reformulations du plan d'études se concentrent sur les contenus, son enseignement et les compétences qu'elle juge nécessaire de développer chez ses élèves.
3. *La prise en compte des processus d'enseignement et d'apprentissage pour penser les pratiques évaluatives.* Cette prise en compte est déjà présente au début de la formation, et elle évolue avec les séminaires. Marquée par un attachement aux moyens, elle s'oriente davantage vers une attention portée aux objets de savoir.
4. *Les connaissances incontournables en évaluation de Samia référées à son contexte.* Si Samia ne semble avoir trouvé avec la taxonomie qu'un outil théorique lui permettant de conscientiser certains implicites « ça avait été un élément important aussi, ne serait-ce que la remise en question de voir que comprendre c'était pas analyser, que c'était... Ça, ça a été... une prise de conscience, je dirais, qui a été importante. Mais... [...] ça me paraissait logique » (entretien 2, 180-181), l'alignement curriculaire élargi, en revanche, nous semble lui avoir permis d'affiner sa connaissance du PER, des caractéristiques des tâches évaluatives, des systèmes de pondération et de notation, au vu de ses propos mais surtout des épreuves créées.
5. *Une pratique évaluative considérée comme relevant, pour Samia, d'un compromis permanent entre différentes dimensions.* Comme Alizée, ces compromis sont très explicites quand Samia décrit le passage entre le travail de critérisation et celui de la

notation comme restant délicat. Là, elle finit par accepter tout phénomène de compensation, et ce malgré l'important travail réalisé en amont. Nous retrouvons d'autres compromis lorsqu'elle reformule, invente ou sélectionne des objectifs par rapport à ceux du PER, tâche ô combien complexe et parfois, chez elle, peu transparente.

6.3 En guise de synthèse provisoire

Au regard de ce qui précède, nous voyons que le développement des compétences évaluatives chez Samia poursuit une logique particulière.

En effet, le fait que d'une part, elle soit encore dans un rapport au PER spontané, et que d'autre part, elle reste encore en questionnement profond sur le lien entre pondération et notation relativement aux différents objets de savoir évalués, nous fait constater qu'aucune compétence ne peut pleinement se développer.

Plus finement, au niveau de la compétence A, c'est la question de la reformulation des objectifs qui demanderait encore à être référée à une taxonomie par exemple. Pour la B, c'est la question de la complexité des tâches, variables au sein d'une même épreuve, qui serait à approfondir. L'impossibilité de poser une évaluation critérielle et de rester attachée aux points affaiblit le développement de la compétence C. Quant à la D, la compensation, indépassable, induit une rupture d'alignement curriculaire élargi au niveau de la notation, ce qui rend l'évaluation moins cohérente (compétence E).

Seule la F, à nos yeux et malgré ce que montre notre tableau, serait significativement en développement. En effet, malgré ces impossibilités, Samia peut justifier la plupart du temps ses choix en matière de pratiques évaluatives sommatives au regard de son contexte institutionnel cantonal, local (établissement) et de la culture des équipes dans lesquelles elles s'inscrivent, en les adaptant aux prescriptions institutionnelles, tout en étant attentive à ce qu'elles soutiennent les apprentissages des élèves.

Ici encore, en référence à notre perspective théorique d'AL, nous observons qu'une tension présente sur une des catégories entraîne de fait une résistance au développement de plusieurs compétences. Il en va ainsi de la question de la pondération et de la notation, dont nous avons déjà parlé. Ce constat revient à nous interroger sur une nouvelle dimension de la pratique évaluative : dans quelle mesure peut-on considérer qu'une évaluation curriculairement alignée

relativement à des objectifs reformulés par l'enseignant peut être cohérente et/ou valide ? Nous y reviendrons dans notre conclusion.

Samia semble donc avoir acquis des connaissances et des compétences clés en s'appropriant le modèle de l'alignement curriculaire. Dans son acception élargie ? Nous serions tentés de répondre par la négative. Nous pensons toutefois qu'une posture réflexive est en voie de développement à certains niveaux grâce au processus de formation, au cœur de laquelle son jugement professionnel a gagné en cohérence grâce à la référence permanente au modèle¹⁶⁷. Sa tentative avortée de système de pondération critérié présent dans sa seconde évaluation en est selon nous un exemple pertinent. Elle met notamment en évidence le fait que certains objets de savoir ne sont pas propices à une telle évaluation, et que le modèle y rencontre donc ses limites. Nous y reviendrons également dans notre partie de discussion-conclusion, ce point nous semblant fort intéressant.

7 Conclusion

Ce chapitre délivre un certain nombre de résultats pertinents à nos yeux, qui apportent non seulement de nouvelles connaissances sur les pratiques évaluatives sommatives, mais qui débouchent aussi sur des questions fortes pour la recherche et pour la pratique. Nous les esquissons ici brièvement, pour en reprendre une partie dans la conclusion finale afin de les développer.

Relativement à notre première question de recherche, il ressort que c'est la notation qui est essentiellement problématique pour chacun des enseignants et que c'est à ce niveau que se situe leur questionnement principal. Il est intéressant de constater qu'il est omniprésent avant la formation, et perdure tout au long de la recherche, sous diverses formes que nous détaillons plus bas. Certes, le rapport au prescrit, et plus particulièrement au plan d'études, est également en tension, mais il ne suscite pas les mêmes interrogations de la part des praticiens. Alors qu'ils « font avec » le PER, ils ne savent régulièrement « pas quoi faire » avec la notation, pris entre des directives souvent décrites comme contraignantes et un désir d'être objectif. Même si certains d'entre eux trouvent des manières de noter innovantes (Alizée et Anouck), d'autres, comme Philémon ou Fernand, restent sur des compromis, et d'autres encore, comme Samia et

¹⁶⁷ Comme nous l'annonçons dans notre partie 6 sur la méthodologie relativement à la recherche collaborative, nous assumons pleinement que la référence au modèle théorique nous intéresse dans une perspective de transformation des pratiques des enseignants.

Paul, semblent ici se fondre dans une posture de renoncement.

C'est donc bien au niveau de la notation que l'existence de ruptures de cohérence d'alignement curriculaire élargi est conscientisée chez tous les enseignants. Et c'est donc à ce niveau que le questionnement s'affine et restera pour chacun en suspens : Comment pratiquer une notation référée à l'apprentissage dans un contexte normatif (Fernand) ? Quelle pratique permettrait de dépasser la limite des points et leur transformation arithmétique en une note (Paul) ? Comment passer de critères à visée de feedback aux élèves à des critères au service de la construction de la note (Philémon) ? Comment construire des seuils de notes pour chacune d'entre elles (Samia) ? Comment transposer une notation référée à l'apprentissage à d'autres contenus que celui de la production écrite (Alizée et Anouck) ? Autant d'interrogations qui nous obligeront à replonger dans nos données dans notre second chapitre de résultats.

Cela étant, nous observons que malgré ces questions ouvertes sur la notation, des mouvements significatifs dans les pratiques évaluatives semblent émerger chez tous les enseignants pour aller vers davantage de cohérence. Nous pensons notamment, et de manière non exhaustive, aux tables de spécification de Fernand pour penser les preuves d'apprentissage dans les tâches évaluatives au regard des objets de savoir et des objectifs d'apprentissages, aux tentatives de Paul de pondérer en référence aux contenus, à la discrimination de la complexité des contenus dans les tâches évaluatives opérée dans les épreuves de Philémon, puis, plus progressivement mais de manière tout aussi significative, la perspective critérielle adoptée par Anouck et Alizée en production écrite, ou encore la réflexion menée par Samia autour des minima attendus pour fixer le seuil de la note 4 pour des contenus fort différents. Ces éléments, selon nous, donnent des indices relativement valides indiquant que le modèle de l'alignement curriculaire élargi, dans une perspective compréhensive, est un outil conceptuel fort porteur pour amener les pratiques évaluatives sommatives des enseignants vers davantage de cohérence, en permettant notamment une meilleure explicitation des apprentissages à mener et à évaluer, un rapport de cohérence plus étroit entre les tâches évaluatives et les objectifs évalués, une plus grande acuité sur la formulation et l'exploitation de critères d'évaluation. Ce résultat est à nos yeux novateur, puisqu'il s'inscrit au niveau de l'étude des pratiques d'évaluation sommative, champ relativement peu exploré dans les recherches portant sur l'alignement curriculaire.

Puis, relativement à notre seconde question de recherche, un état de fait ressort : tous les enseignants ont opéré des changements importants dans leurs pratiques évaluatives sommatives suite au premier séminaire de formation lors duquel ils ont été amenés à s'approprier le modèle

théorique de l'alignement curriculaire élargi, puis à analyser et concevoir une épreuve sommative à l'aune de ce dernier. Ces changements - que nous concevons davantage comme des mouvements jamais stabilisés - sont non seulement observables chez tous les participants, mais ont aussi mis en évidence des réflexions nouvelles, comme si le modèle permettait une explicitation d'interrogations ancrées souvent dans des pratiques peu questionnées. Ces réflexions - et c'est ici que l'appropriation du modèle interpelle - ont eu des impacts sur les pratiques rapides et marquants, et corollairement, sur le développement des compétences évaluatives des six enseignants. Nous l'avons montré, ce développement a beau être systémique, donc forcément irrégulier, non linéaire et dépendant de multiples éléments contextuels, il n'en reste pas moins, dans notre recherche, également interpellant. En effet, nous observons qu'il existe une corrélation entre l'appropriation du modèle théorique par les enseignants comme outil compréhensif de leurs pratiques évaluatives sommatives et le développement des six compétences en jeu dans la recherche. À ce titre, nous observons que les compétences A et B, relatives à l'exploitation du plan d'études et au *design* des tâches, se développent significativement chez tous les participants. La compétence C, qui aborde la capacité à élaborer des systèmes de pondération complexes, se développe essentiellement chez Fernand, Alizée et Anouck, qui sont les seuls à avoir soit construit des tables de spécification, soit engagé des démarches de notation critérielles. En revanche, c'est au niveau des compétences D et E, portant sur la notation et la justification de la cohérence des pratiques, que des différences de développement s'observent tout particulièrement. Alors qu'Alizée et Anouck font preuve encore une fois de connaissances mises en pratique dans leurs épreuves et peuvent les justifier, leurs collègues, pour différentes raisons, ne peuvent ici aller de l'avant. Il en va de même pour la compétence F, où l'on observe qu'une partie des enseignants rencontre de forts dilemmes lorsqu'ils tentent de faire cohabiter des pratiques évaluatives sommatives innovantes et des contraintes contextuelles ressenties comme (trop) fortes. Anouck et Alizée, qui ont mis en œuvre des pratiques évaluatives critérielles jusqu'à la notation, sont les participantes qui semblent le plus avoir développé leurs six compétences. À l'inverse, Samia, dont le rapport au PER semble rester peu outillé et les réticences à envisager des critères jusqu'à la notation persister, est celle dont les compétences paraissent le moins se développer. Entre ces deux « extrêmes », les développements ne sont jamais identiques et soulignent la dimension dynamique de la perspective théorique de l'*assessment literacy*.

Ces constats ouvrent pour nous un questionnement conceptuel relatif à l'articulation de notre modèle théorique avec les pratiques réelles. Nous le commençons ici, mais une discussion plus

approfondie que nous mènerons dans notre chapitre conclusif va s'avérer nécessaire. En effet, au regard de ces premiers résultats, nous pouvons nous demander si c'est au travers de l'appropriation du modèle de l'alignement curriculaire que les enseignants ont pu interroger, voire infléchir leurs pratiques évaluatives sommatives, ou de son acception élargie ? La nuance, qui semble anecdotique, est de taille. Conceptuellement d'abord, puisque l'enjeu théorique est ici de saisir à quelles conditions le modèle peut permettre une meilleure compréhension des pratiques évaluatives sommatives notées. Mais aussi, et peut-être davantage, d'un point de vue praxéologique, puisque, dans un dispositif de recherche collaborative comme le nôtre, nous poursuivons une recherche d'optimisation de l'action des enseignants et tentons de comprendre à quelles conditions le modèle peut être saisi dans les pratiques évaluatives. Cette nuance peut se traduire par la question suivante : un enseignant peut-il être un bon évaluateur et un « notateur » peu cohérent, voire discriminant ? En d'autres termes, peut-on concevoir qu'un enseignant développe des pratiques alignées curriculairement, mais désalignées d'un point de vue élargi ? Même si nous sommes tentés de répondre ici par l'affirmative, ces questions nous obligeront, également dans notre conclusion, à étayer le rapport entre évaluation et notation, puisque certains des enseignants de notre recherche vont vers plus de cohérence dans leurs pratiques de *design* ou de référencement restreinte, puis, pris dans diverses contraintes lorsqu'il faut noter, semblent « décrocher » de leur recherche de cohérence. Certes, l'appréhension de cet obstacle au niveau de la notation par les enseignants est moins un aveu d'impuissance qu'une impossibilité, souvent compréhensible et justifiée, de mettre en œuvre une vision particulière de l'évaluation sommative inscrite dans le modèle : compréhensive, qualitative, au service des apprentissages. Le contexte, à ce niveau, a sa part de responsabilités dans les empêchements décrits par les enseignants, et leur bilan reste somme toute positif. Mais il reste que l'acception élargie du modèle semble se confronter à des contraintes empêchant de concevoir des épreuves qualitatives jusqu'à la notation.

Enfin, précisons que si ces premiers résultats sont porteurs de nouvelles connaissances sur les pratiques évaluatives sommatives appréhendées à l'aune d'un modèle théorique, nous pensons toutefois qu'il serait abusif de les considérer comme définitifs, notre perspective de recherche étant éminemment contextuelle et compréhensive. Dans ce sens, nous touchons, avec cette réflexion, une des limites de notre étude dont nous débattons également dans notre conclusion.

Partie 8 - Second chapitre de résultats

1 Introduction

Les résultats de notre chapitre précédent ont mis en évidence la nécessité d'investiguer plus profondément les pratiques de notation. D'une part parce que c'est à ce niveau que les enseignants ont particulièrement porté leur questionnement du début à la fin de la recherche, et d'autre part car elles entretiennent, avec le modèle théorique de l'alignement curriculaire élargi, un rapport complexe qui peut à certaines conditions aboutir sur des transformations possibles de pratiques, mais aussi régulièrement sur des tensions voire des résistances.

C'est au travers de la question générale de recherche suivante que nous analyserons nos données : *Comment les enseignants décrivent-ils leurs pratiques de notation ?* Le but ici est de donner à voir sur quels phénomènes relatifs à la notation les enseignants portent leur attention, soit pour manifester des problèmes rencontrés, soit pour expliciter des questions professionnelles nécessitant des apports de la recherche. Cette question générale se divise en trois sous-questions :

1. *Sur quel.s phénomène.s, question.s et réflexions les enseignants portent-ils leur attention relativement à leurs pratiques de notation avant la formation ?* Il nous a semblé ici important de recenser les dimensions sur lesquelles les participants sont en questionnement quand ils notent avant de débiter le processus de formation continue proposé, alors qu'ils ne connaissent pas le modèle théorique de l'alignement curriculaire élargi, et qu'ils sont pris dans leurs habitudes, leurs contraintes, et des directives qu'ils suivent parfois machinalement alors qu'ils n'en perçoivent ni l'origine ni le sens. Ce questionnement nous semble fécond d'un point de vue de formation continue car il met en exergue des « questions vives de la pratique », c'est-à-dire des questions qui engagent les enseignants dans un processus de réflexion porté par leurs soins. Cette richesse est également intéressante pour la recherche, dans le sens où elle s'inscrit dans le sillage de travaux antérieurs sur les pratiques évaluatives sommatives (e. g., Mottier Lopez, Tessaro, Dechamboux & Morales Villabona, 2012) et en propose des prolongements possibles.
2. *Quelle est la nature des obstacles rencontrés par les enseignants dans leurs pratiques de notation ?* Nous l'avons montré, si les pratiques de notation de la plupart des enseignants, globalement, gagnent en cohérence, elles font également face à un certain nombre d'obstacles qu'il s'agit de documenter. Nous tenterons ici de montrer le rôle des

outils d'une part, dans les tensions voire impossibilités qu'ils génèrent, mais aussi le rôle des disciplines, voire des objets disciplinaires. Nous souhaitons comprendre pourquoi, par exemple, Alizée et Anouck ont réussi à développer des pratiques de notation cohérentes d'un point de vue d'alignement curriculaire élargi dans des épreuves portant sur la production écrite en français et ne se sont pas aventurées sur un autre objet de savoir, ou, corollairement, pour quelles raisons aucun enseignant de mathématiques n'y est parvenu dans le cadre de cette recherche.

3. *Quels apports sur les pratiques de notation ces résultats offrent-ils du point de vue de l'alignement curriculaire « élargi » ?* En répondant à cette question, nous tenterons de mettre en évidence des éléments de compréhension à deux niveaux :

- Un niveau théorique, de manière à affiner conceptuellement les rapports entre évaluation et notation. En effet, comme nous l'avons proposé dans notre partie conceptuelle, leur relation doit encore être davantage soumise à des données de recherche afin d'être encore plus finement thématisée ;
- Un niveau praxéologique relatif au développement d'une littératie en évaluation (*assessment literacy*) pour éprouver dans quelle mesure la notation – et les connaissances qu'elle requiert – est une dimension clé de toute compétence évaluative sommative.

Notons que fidèlement à la logique sous-tendue par notre modèle théorique, nous serons amenés à considérer que certains phénomènes relevés par les enseignants au niveau de la pondération seront également à mettre en exergue ici. Comme nous l'avons mis en évidence dans notre partie précédente, les pratiques de pondération infléchissent les pratiques de notation, même si elles n'en prédéfinissent pas les caractéristiques une fois pour toutes. Par exemple, nous avons vu les enseignants de mathématiques adopter souvent des critères qualitatifs sans pour autant les exploiter dans des échelles descriptives. Dès lors, nous faisons l'hypothèse que d'un point de vue de recherche mais aussi de formation, des questions relevant de la notation peuvent trouver une partie de leur réponse en investiguant les pratiques de pondération.

Conjointement, nous aborderons les questions des outils créés par les enseignants pour noter. Cela nous amènera spécifiquement à aborder la question des échelles de notation.

2 La description des pratiques de notation par les enseignants

Commençons par un état des lieux avant la formation.

2.1 Les phénomènes, questions et réflexions sur lesquels les enseignants portent leur attention dans leurs pratiques de notation avant la formation

Nous répondrons à cette sous-question en structurant notre propos en trois temps correspondant chacun à un tableau d'organisation des données suivi de commentaires interprétatifs.

- Le premier tableau montre sur quel phénomène relatif aux pratiques de notation chaque enseignant porte son attention et met en évidence le questionnement suscité. De manière à rester fidèle à notre organisation, nous groupons les enseignants par discipline, en commençant par les enseignants de mathématiques.
- Le deuxième tableau met en correspondance chaque phénomène avec les enseignants qui l'ont identifié comme étant problématique ou porteur de questions/réflexions. Ce mode de présentation nous permettra d'esquisser des pistes d'interprétation quant à la récurrence de certains phénomènes.
- Enfin, le troisième tableau offre une entrée disciplinaire, afin de voir dans quelle mesure cette dernière peut constituer une perspective interprétative cohérente pour comprendre l'émergence des phénomènes relevés.

Tableau 23 : Phénomènes, questions et réflexions mis en évidence par les enseignants sur leurs pratiques de notation avant la formation

Enseignant.e	Phénomène relevé	Questions et réflexions y relatives	Discipline
Fernand	L'attribution du poids à l'apprentissage évalué (pondération)	Dans une évaluation, quel poids donner à chaque partie, à chaque objectif ?	Mathématiques
	Le manque de sens des barèmes standardisés	Utilisation univoque du barème fédéral	
Paul	Le barème standardisé fausse l'orientation	L'exploitation systématique du barème fédéral ne permet pas de noter correctement ; il est à la base d'une orientation scolaire souvent peu étayée	
	Le manque de justesse du barème	Changement des seuils a posteriori parfois	
		La moyenne de classe comme référence de la réussite	
Philémon	La compensation	Ouverture à poser des points ou des demi-points qui se compensent	
	L'attribution du poids à l'apprentissage évalué (pondération)	Evalue-t-on les nouvelles connaissances ou tout le temps la même chose ?	
	Le décalage entre les acquis des élèves et les notes attribuées	Peut-on avoir de bonnes notes avec des réponses erronées ?	
	L'établissement du seuil de suffisance	En fonction de ce que les élèves doivent maîtriser	

	La linéarité des échelles	La linéarité est un piège !	
Anouck	La justesse de la note	Comment savoir que la note reflète les compétences de l'élève ?	Français
	La construction du seuil de suffisance	Au travers d'un comptage de points minimaux que les élèves devraient avoir	
	Le rôle prépondérant de l'échelle	C'est l'échelle qui va poser la note plus que l'épreuve	
	Le moment de création de l'échelle	Pas de règle : parfois avant, parfois après, parfois avant puis modification	
	L'adaptation des barèmes	Nécessité de faire correspondre l'échelle à la difficulté de l'évaluation	
	La régularité des échelles comme preuve de justesse	Incapacité à créer des échelles irrégulières	
Alizée	La nature de l'échelle	Comment et quand fixe-t-on les échelles ? Comment choisir les pourcentages de points pour les seuils ?	
		Attrait pour les échelles régulières, sans justification	
		Le barème fédéral n'a pas de sens	
	Le moment de construction de l'échelle	Doit-il intervenir avant ou après l'évaluation ?	
	La construction du seuil de suffisance	Au travers d'un pourcentage de points, mais le calcul est-il juste en fonction de l'épreuve ?	
Samia	Le sens de la note 4	En production écrite, comment noter en incluant du fonctionnement de la langue ?	
	Le flou des échelles et des barèmes	Est-ce que c'est chercher trop loin que de calculer le seuil du 4 puis d'utiliser « ibaremes » pour des échelles régulières ?	

Ce tableau nous montre que les six enseignants pointent la notation comme source de réflexion ou de questionnement profond avant la formation pas moins de vingt fois, ce qui est conséquent au regard des autres dimensions de la pratiques évaluative investiguée à ce moment-là de la recherche.

Ainsi, nous observons que chez les spécialistes de mathématiques, Fernand s'interroge sur ses pratiques de pondération, car il n'est pas clair pour lui de savoir quel poids donner à l'apprentissage évalué, et surtout, comment signifier cette répartition de poids (par objectif, par tâche ?). Le rapport entre la pondération et la notation est aussi au centre de ses préoccupations. Là, il n'est pas satisfait des barèmes standardisés qu'il exploite car ils n'ont pour lui guère de sens. Paul est également en questionnement sur les mêmes objets. En effet, il pratique la compensation des points lorsqu'il pondère, et croit en la moyenne de classe comme référence pour évaluer la qualité de son enseignement. Cela étant, il discrédite le barème fédéral qu'il utilise par ailleurs, car il est conscient qu'en le modifiant, il peut faire dire ce qu'il veut à une note. Philémon, quant à lui, semble plus critique sur ces mêmes questions, puisqu'il reconnaît

être prisonnier de la linéarité des échelles qu'il utilise. C'est hypothétiquement ce questionnement qui l'empêche de comprendre le décalage important qu'il rencontre souvent entre les notes qu'il administre aux élèves et leurs réelles acquisitions relevées dans les épreuves, et cela, paradoxalement, alors qu'il élabore ses seuils de suffisance au regard des contenus à maîtriser, ce qui devrait être un gage de cohérence.

Du côté des enseignantes de français, nous observons qu'Anouck est méfiante vis-à-vis de la note, car il n'est jamais évident de savoir dans quelle mesure elle rend compte des acquis des élèves. Si elle fait état de pratiques peu questionnées (élaboration du seuil de la note 4 en comptabilisant un nombre minimal de points, rapport étroit entre la complexité de l'épreuve et l'échelle, régularité des échelles), elle oscille sur le moment d'élaboration de l'échelle, pas défini. C'est par ailleurs cette même échelle qui semble bien davantage détenir pour elle le pouvoir d'établir la note que l'épreuve. En revanche, Alizée, sa collègue directe, est quant à elle en grand questionnement sur la notation. Elle se demande comment et quand construire une échelle, quel rapport entretiennent les seuils des notes avec l'épreuve. Mais elle est en même temps séduite par les échelles régulières tout en étant critique face à la pertinence du barème fédéral, sans être toutefois en mesure de justifier ses positions. Enfin, Samia, également prise dans cette question de régularité, semble très peu sûre de ses pratiques de notation. Elle les questionne au regard des contenus auxquels les notes renvoient et à leur mode de structuration. Et les confronte à des situations singulières, comme lorsqu'elle est amenée à noter des productions écrites dans lesquelles figurent des éléments du fonctionnement de la langue. Là, elle ne sait comment faire.

Individuellement donc, ce qui frappe ici, c'est que chacun des six enseignants explicite des pratiques de notation peu outillées, relativement ritualisées, mais surtout en questionnement. Il serait selon nous toutefois réducteur d'en rester là. En effet, une description plus poussée des causes de ces questions et tensions doit avoir lieu. C'est la raison pour laquelle nous présentons, dans le tableau suivant, la fréquence de chaque phénomène et des questions/réflexions y relatives en les mettant en correspondance avec les enseignants qui les ont relevés. Cela nous permettra une lecture croisée sur ce qui émerge avant la formation.

Tableau 24 : Types de phénomènes, questions et réflexions mis en évidence par les enseignants sur leurs pratiques de notation avant la formation

Phénomène mis en question, interrogations et réflexions	Enseignant	Discipline
L'attribution du poids à l'apprentissage évalué et les questions de compensation	Fernand	Mathématiques
	Paul	
	Philémon	
Le manque de sens des barèmes standardisés	Fernand	Mathématiques
	Paul	
	Philémon	
	Anouck	Français
Le manque de justesse ¹⁶⁸ du barème	Paul	Mathématiques
	Philémon	
	Anouck	Français
	Samia	
L'établissement du seuil de suffisance	Philémon	Mathématiques
	Anouck	Français
	Alizée	
La linéarité des échelles	Philémon	Mathématiques
	Anouck	Français
Le rôle prépondérant de l'échelle	Anouck	Français
Le moment de création de l'échelle	Anouck	Français
	Alizée	

De manière à faciliter la lecture des phénomènes relevés et éviter les redondances, nous en avons regroupés certains. Ainsi, il nous a semblé pertinent de considérer que *l'attribution du poids à l'apprentissage évalué et les questions de compensation* relevaient ici d'un phénomène identique, celui de la pondération dans une visée de notation. Dans la même logique, nous avons considéré, au regard des résultats précédents, que lorsque Paul dit que les barèmes standardisés faussent l'orientation, il questionne leur *justesse*. Idem pour Philémon, quand il relève le décalage entre les acquis des élèves et les notes attribuées, ou pour Samia qui condamne le flou des échelles et des barèmes.

Il ressort donc de ce deuxième tableau les constats suivants :

- Le phénomène « *attribution du poids à l'apprentissage évalué et questions de compensation* » est un phénomène essentiellement relevé par les enseignants de mathématiques.
- *Le manque de sens des barèmes standardisés* est également relevé par les trois mêmes spécialistes. Seule Anouck, en français, partage cette préoccupation.
- *Le manque de justesse du barème* est relevé par deux enseignants de chaque discipline.

Il est intéressant de noter ici que Paul et Philémon questionnent conjointement le sens

¹⁶⁸ Cette justesse renvoie au souci des enseignants de produire des notes, selon leurs dires, « justes », c'est-à-dire des notes qui rendent compte le plus objectivement possible de la qualité des productions évaluées.

et la justesse de cet outil, comme Anouck en français, alors que Fernand dissocie sens et justesse, comme Samia.

- *L'établissement du seuil de suffisance* n'est abordé que par trois participants. Alors que Philémon et Anouck le calculent en fonction de ce que les élèves doivent maîtriser dans les épreuves, Alizée questionne sa construction.
- *La linéarité des échelles*, quant à elle, divise. Philémon s'y sent prisonnier et Anouck se dit incapable d'en sortir, sans pouvoir se justifier. Peut-être parce qu'elle considère le rôle de l'échelle plus important que l'épreuve en elle-même ?
- Enfin, *le moment de création de l'échelle* est assez flou chez Alizée et Anouck, les deux seules qui en parlent. Avant ou après la passation de l'évaluation ? Rien n'est clair.

Deux niveaux de questions/réflexions se côtoient selon nous ici et permettent une interprétation des données. Un premier niveau de compréhension général donne à voir des enseignants en difficultés pour comprendre l'exploitation et la construction de barèmes standardisés et en expliciter les mode opératoires (comme la linéarité). Un second niveau relatif à la pratique de notation est conjointement observable, dans lequel on peut voir les praticiens se poser des questions de l'ordre du « comment faire » : sur quelles bases construire les seuils des notes ? À quel moment élaborer les échelles ?

Ce constat tendrait à montrer que l'élaboration d'outils à visée de notation est très complexe, puisqu'elle suscite un questionnement tout aussi bien idéologique que pratique. Mais qu'en est-il du rôle de la discipline dans un tel questionnement ?

Tableau 25 : Par discipline, phénomènes, questions et réflexions mis en évidence par les enseignants sur leurs pratiques de notation avant la formation

Discipline	Enseignants	Phénomène mis en question, interrogations et réflexions
Mathématiques	Fernand, Paul et Philémon	L'attribution du poids à l'apprentissage évalué et les questions de compensation
	Fernand, Paul et Philémon	Le manque de sens des barèmes standardisés
	Paul et Philémon	Le manque de justesse du barème
	Philémon	L'établissement du seuil de suffisance
	Philémon	La linéarité des échelles
Français	Anouck	Le manque de sens des barèmes standardisés

	Samia et Anouck	Le manque de justesse du barème
	Anouck et Alizée	L'établissement du seuil de suffisance
	Anouck	La linéarité des échelles
	Anouck	Le rôle prépondérant de l'échelle
	Anouck et Alizée	Le moment de création de l'échelle

Force est de constater ici qu'il ne semble pas y avoir de phénomène relatif à la notation spécifique à une des deux disciplines chez les enseignants avant la formation. Chaque phénomène, peu ou prou, est relevé par l'un ou l'autre d'entre eux. C'est donc vraisemblablement face à un questionnement bien transversal que nous nous trouvons, même si Anouck est la seule à relever le rôle prépondérant de l'échelle, ce qui ne constitue pas en soi un résultat significatif.

2.1.1 Quelques éléments sur le rapport à la notation des enseignants avant la formation

Nos analyses croisées au travers des trois tableaux confirment le fait que la notation est une pratique, voire même un processus, qui questionne profondément les six enseignants avant d'entrer en formation. Nous serions tenté de dire que pour aucun d'entre eux, cette phase de l'évaluation sommative n'est ni anodine, ni basement technique. Au contraire. Nous observons que derrière les phénomènes relevés se cachent des questions vives de la pratique, et que bon nombre de ces questions concernent les échelles. Parfois même, les interrogations sont idéologiques, empreintes de valeurs, et c'est davantage la posture de l'évaluateur qui est ici selon nous au cœur des interrogations.

Nous avons vu dans la partie précédente que la majorité des enseignants ont opéré des mouvements dans leurs pratiques de notation. Mais des obstacles perdurent. Il s'agit maintenant d'investiguer leur nature.

2.2 Nature des obstacles rencontrés par les enseignants dans leurs pratiques de notation

Au travers de notre précédente question de recherche, nous avons présenté l'état des lieux des pratiques de notation des participants et de leurs questions/réflexions y relatives avant la formation. Dans la partie précédente, nous avons mis en évidence le cheminement de chacun au niveau de ses pratiques évaluatives. Nous nous arrêtons ici à la fin de la recherche collaborative, pour dresser un état des lieux des pratiques de notation de chacun d'entre eux. Le but est de rendre visible les mouvements significatifs à ce niveau, mais surtout de mettre en discussion les obstacles persistants, afin de mieux les comprendre. Le tableau ci-dessous tente mettre ces deux dimensions en dialogue pour mieux saisir la complexité des difficultés

rencontrées par les praticiens.

Tableau 26 : Mouvements opérés dans les pratiques de notation et nature des obstacles persistants en fin de formation

Enseignant	Mouvements opérés pour une notation cohérente d'un point de vue d'alignement curriculaire élargi	Obstacles persistants dans les pratiques de notation pour tendre vers cette cohérence	Discipline
Fernand	Les pratiques de notation sont référées aux apprentissages, basées sur des points liés à des critères. Elles relèvent également de compromis. Une importance particulière est accordée au seuil du 4 qui est aligné curriculairement, le plus possible référé à des unités d'apprentissage	La construction des notes « écrase » toujours le travail qualitatif réalisé en amont par l'utilisation d'échelles linéaires, sauf pour le 4, le plus important. Décrire chaque note est subjectif, le barème fédéral est alors rassurant	Mathématiques
Paul	Il y a un sens à tendre vers un jugement référé à des unités d'apprentissage pour donner du feedback aux élèves, mais sans aller jusqu'au verdict sommatif. Le seuil du 4 est élaboré au regard des minima attendus	Le barème fédéral reste la référence pour construire les échelles, mieux adapté à une pondération en points, et malgré le fait qu'il soit remis en question dès le début de la formation. Le recours à une échelle arbitraire est toujours présent. Le barème standardisé reste la référence. La distribution des résultats comme gage de qualité de l'évaluation est toujours présente	
Philémon	Les seuils sont pensés en références aux attentes liées à l'apprentissage, en amont, notamment celui du 4. La notation est en rapport étroit avec la pondération	La compensation est un phénomène accepté et la notation référée à l'apprentissage est problématique, notamment à cause de la multitude des échelons (onze)	
Anouck	Une notation référée à l'apprentissage semble se stabiliser, sauf sur certains contenus, comme le fonctionnement de la langue	Des dilemmes persistent sur la linéarité à construire avec des grilles critérielles. Dans une perspective qualitative, difficulté à passer d'une note à l'autre (différence entre un 4 et un 4,5 ?)	Français
Alizée	Les notes ont gagné en cohérence : Une autre manière de fixer le 4 est à l'oeuvre, en référence au contenu. Il n'est pas pertinent d'aller en dessous de la note 3	La nécessité d'expérimenter ce type de démarche de notation sur d'autres objets de savoir est présente	
Samia	Le raisonnement pour aboutir à ce 66% (seuil du 4) est complètement différent. La notation est pensée en fonction de ce qui est maîtrisé. Le barème en amont permet de repenser les tâches évaluatives	Une attention à une forme de linéarité des échelles, appelée « équité », est encore présente	

Trois obstacles majeurs allant à l'encontre de pratiques de notations alignées d'un point de vue curriculaire nous semblent émerger de nos données.

2.2.1 L'obstacle du barème et des échelles

Tous les enseignants mentionnent cet obstacle. Il est frappant de noter que malgré des mouvements de pratiques évaluatives impactant positivement leur cohérence chez chacun d'entre eux, le barème reste problématique à exploiter, et, plus interpellant, à construire, surtout lorsqu'une recherche de notation référée à l'apprentissage est à l'œuvre. À ce titre, nous observons que le recours à des barèmes standardisés, issus d'algorithmes facilement reproductibles (comme le barème fédéral), est vécu comme rassurant, voire comme une issue de secours acceptable lorsque l'enseignant doit prendre en main la construction de la note. Ces décisions sont parfois prises avec une forme de résignation (Fernand, Philémon, Samia), mais aussi avec une certaine conviction (Paul), ce qui tendrait à confirmer que le problème dans la notation est moins son existence que son mode de construction. Nous l'avons dit dès notre introduction, les échelles imposent aux enseignants vaudois d'exploiter pas moins de onze échelons (notes de 1 à 6, avec demi-notes). Et c'est bien cet empan qu'ils ont de la difficulté à traduire en autant de progressions dans la qualité de l'apprentissage évalué. L'hypothèse que nous formulons dès lors ici est la suivante : avec des échelles plus réduites, ou dans lesquelles l'insuffisance ne serait pas signifiée au travers de six notes, mais d'une ou de deux¹⁶⁹, la possibilité de construire des notes référées à l'apprentissage et donc cohérentes d'un point de vue d'alignement curriculaire élargi serait bien davantage opérationnalisable. Certains enseignants prennent d'ailleurs ce chemin dans leurs réflexions et/ou pratiques, en réduisant leurs échelles (de 2.5/3 à 6 pour Alizée et Anouck), en envisageant d'abandonner les demi-notes (Philémon), ou en faisant en sorte que les notes entre 1 et 2.5 soient quasi impossibles à obtenir par des élèves en difficulté (Paul, Philémon, Fernand, Samia)¹⁷⁰.

Cet attachement aux barèmes et échelles standardisés n'est par ailleurs pas sans lien selon nous avec la relation complexe que certains enseignants continuent à avoir avec la linéarité possible de ces outils. Là, nous observons qu'elle peut être ici considérée comme un gage d'équité dans la construction de la note (Samia), mais qu'elle interroge aussi ce qui la constitue. En effet, d'un point de vue pratique, qu'est-ce qu'une linéarité d'échelle ? Si l'on peut bien la concevoir d'un point de vue mathématique lorsque cette dernière est construite sur la base de points, qu'en

¹⁶⁹ Comme c'est le cas en Finlande au niveau secondaire, où les notes vont de 4 à 10 et où l'insuffisance n'est signifiée qu'au travers de la note 4.

¹⁷⁰ Les stratégies en jeu relèvent de pratiques de pondération et de *design* modélisées au travers de l'alignement curriculaire élargi que nous présentons en détail dans la partie précédente. En résumé, elles consistent à construire des seuils d'insuffisance grave très bas, ou, conjointement, à donner l'occasion aux élèves d'engranger des points ou de valider des critères sur des objets de savoir clés en y insistant dans l'enseignement.

est-il lorsque l'échelle devient descriptive dans une perspective pédagogique ? En d'autres termes, une linéarité référée à l'apprentissage peut-elle être concevable, et si oui, à quelles conditions ? Ce questionnement, très pertinent à nos yeux, est totalement assumé par Alizée et Anouck dans leurs élaborations d'échelles descriptives¹⁷¹, et nous serions tentés de dire qu'il en est en partie de même pour Paul et Fernand, puisque, s'ils ne vont pas jusqu'à noter en référence à des unités d'apprentissage, leurs pratiques de pondérations critériées les aident tout de même à construire des progressions dans leurs attentes. Mais ce questionnement met aussi à mal les autres collègues, pour des raisons vraisemblablement différentes. Un constat qui nous amène à traiter du deuxième obstacle.

2.2.2 L'obstacle des seuils

Nous pourrions penser que l'obstacle relevant de l'utilisation et de la construction des barèmes a directement un rapport avec celui relatif à l'élaboration des seuils des notes, mais force est de constater que ce lien est à nuancer. Tous les enseignants montrent qu'ils pensent et construisent le seuil de suffisance traduit par la note 4 de manière complètement différente à la fin de la formation, puisqu'il est dorénavant référé à des apprentissages minimalement attendus. C'est un résultat significatif de notre recherche. Toutefois, c'est la détermination des autres seuils¹⁷² qui pose encore problème chez une majorité d'enseignants, surtout ceux qui concernent les notes allant de 4 à 6, ou pour le dire autrement, de la suffisance à l'excellence. En effet, à l'instar d'Anouck ou de Fernand, on pourra se demander ce qui différencie un 4 d'un 4.5, ou, plus délicat encore, un 5.5 d'un 6, quand ces interrogations cherchent des réponses en termes d'apprentissage et non de totaux de points ou de pourcentages. Il semble donc relativement aisé, pour les enseignants, de passer ici d'un questionnement qui viserait à savoir à quel pourcentage il faut poser le 4 à un questionnement cherchant à déterminer ce que l'élève doit minimalement maîtriser pour obtenir cette note. Mais le questionnement, sauf chez Alizée et Anouck, s'arrête là, devenant fort délicat pour discriminer le « très bon » de l'« excellent », voire l'« insuffisant » du « très insuffisant », en d'autres termes, quand il s'agit de mettre en correspondance la note chiffrée avec des catégories d'appréciations qualitatives.

Nous faisons ici l'hypothèse que ce malaise à ne pas parvenir à faire ce travail de traduction est la raison pour laquelle, faute de mieux, les enseignants ont finalement recours aux barèmes

¹⁷¹ Voir annexes Partie 4, 1.2.1 et 1.2.2.

¹⁷² En nous référant à des apports théoriques comme ceux relatifs à la notation holistique (Royce Sadler, 2009), nous parlons de seuil pour l'établissement de chaque note, et non plus seulement du seuil de suffisance (note 4).

standardisés. Et nous pourrions affirmer qu’être compétent à construire un seuil de suffisance est déjà un pas important et fondamental dans des pratiques de notation alignées curriculairement, puisque c’est sur cette base que l’enseignant peut déterminer si l’élève « sait » ou « ne sait pas », « réussit » ou « échoue ».

Ce résultat reste toutefois questionnant d’un point de vue de recherche, puisqu’il montre que le contexte de production de la note et la nature des outils imposés aux enseignants résistent au pouvoir explicatif du modèle théorique de l’alignement curriculaire élargi, ou corollairement, que le modèle ne parvient pas à résoudre ce désalignement curriculaire. En effet, si tous les enseignants semblent comprendre comment potentiellement construire une note idéalement référée à l’apprentissage, seuls deux y parviennent vraiment. Ce qui nous fait dire que ce résultat est également frustrant d’un point de vue de formation. Dans un système scolaire où les notes déterminent non seulement les moyennes de disciplines, mais entrent également en jeu dans le calcul de moyennes de groupes de disciplines qui déterminent l’orientation de l’élève dans ses trois dernières années de scolarité¹⁷³, nous questionnons le fait qu’un tel raisonnement sur les notes allant de 4.5 à 6 ne soit envisageable qu’à certaines conditions, et au prix d’un effort de formation conséquent et cohérent du point de vue théorique. En effet, nous pensons que les enseignants devraient aussi être compétents à construire les notes qui vont permettre aux élèves d’être orientés dans les filières d’excellence. Sans parler du contexte d’enseignement qui joue un rôle important et qui peut, comme pour Fernand, totalement brider les initiatives allant vers une recherche de cohérence des pratiques de notation.

Cela étant, nous avons jusqu’ici traité les obstacles de manière transversales, mais certaines données nous poussent à en considérer un dernier, relatif aux disciplines, qui pourrait encore davantage éclairer certaines interrogations encore ouvertes.

2.2.3 L’obstacle de l’objet de savoir évalué

Pourquoi Alizée et Anouck sont-elles les seules à avoir pu élaborer des notes au travers d’échelles descriptives ? Comment se fait-il qu’elles restent en questionnement sur les possibles transferts de telles pratiques sur d’autres contenus ? Pour quelles raisons Samia s’est-elle acharnée, jusqu’à en être découragée, à tenter de critériar puis noter une épreuve portant sur le

¹⁷³ Comme le montre le document de l’annexe Partie 2, 1.

fonctionnement de la langue, sans succès¹⁷⁴ ? Comment comprendre que les trois enseignants de mathématiques, sur des objets de savoir différents, ne sont pas parvenus à construire des notes en référence à leur immense travail de recherche de cohérence curriculaire réalisé en amont ? Autant de questions qui nous font faire l'hypothèse suivante : un des obstacles majeurs à l'élaboration d'une note référée à l'apprentissage est l'objet de savoir évalué.

Le tableau ci-dessous rend compte des objets de savoir sur lesquels ont porté les dernières évaluations créés par les enseignants et des systèmes de pondération et de notation exploités¹⁷⁵. Cette recension nous permet ensuite de les mettre en perspective au regard des obstacles spécifiques persistants.

Tableau 27 : mise en perspective des objets de savoir évalués dans la dernière épreuve avec les systèmes de pondération, la notation et les obstacles rencontrés dans les pratiques

Discipline	Enseignant	Objet de savoir évalué dans la dernière épreuve	Système de pondération exploité	Notation	Obstacle rencontré dans la pratique de notation
Mathématiques	Fernand ¹⁷⁶	Aires et périmètres	Total de points par progression d'apprentissage du PER, et par exercice, de manière à pouvoir établir des seuils pour les différentes notes	Echelle fixée en amont, calcul du seuil du 4 en fonction des minima attendus	Difficulté à décrire les autres notes. Utilisation du barème fédéral pour se rassurer. Conscience que cette procédure écrase le travail qualitatif réalisé en amont

¹⁷⁴ Voir annexe Partie 5, 9.2.4.

¹⁷⁵ Comme au point 2.1, nous parlons également de pondération puisque cette composante, fidèlement au modèle de l'alignement curriculaire élargi, est en lien dialectique avec la notation et la conditionne. Pour les détails de ces démarches et l'analyse de ces épreuves, le lecteur se référera à la partie précédente, ou à la conclusion de ce manuscrit.

¹⁷⁶ Voir annexe Partie 5, 9.1.2.

Français	Paul ¹⁷⁷	Fonctions et algèbre : calcul littéral	Deux logiques : une référée l'apprentissage et permettant de voir les seuils de suffisance par élément de progression des apprentissages du PER, l'autre plus traditionnelle par l'exploitation de points	Deux logiques : une en rapport avec le système de pondération en points en lien avec les progressions d'apprentissage du PER et l'autre à l'aide du barème fédéral	Le système d'appréciation « insuffisant (1)-suffisant (2)-complètement acquis (3) » par progression d'apprentissage du PER reste difficile à appliquer
	Philémon ¹⁷⁸	Transformations géométriques (symétries et homothéties)	Total de points	Seuils pensés en référence aux attentes liées à l'apprentissage, en amont, notamment celui du 4. Echelle linéaire avant et après le 4	Compensation possible. Difficulté à qualifier les onze notes. Difficulté à sortir de la linéarité de l'échelle
	Alizée ¹⁷⁹	Production écrite (le texte explicatif)	Critères qualitatifs référés à des unités d'apprentissage	Echelle descriptive	Difficulté à envisager un tel système pour d'autres contenus comme le fonctionnement de la langue
	Anouck ¹⁸⁰	Production écrite (le texte explicatif)	Critères qualitatifs	Echelle descriptive	Le barème, après utilisation, ne donne pas satisfaction
	Samia ¹⁸¹	Conjugaison : subjonctif présent et passé, impératif présent et passé	Points référés aux minima attendus dans chaque exercice, puis total de points indépendant de ces attentes	Seuil du 4 calculé à l'aide du total de points minimalement attendu dans chaque exercice	Compensation possible. Impossibilité d'entrevoir des pratiques permettant de passer de critères à une note sur une échelle à onze positions

Pour les mathématiques

¹⁷⁷ Voir annexe Partie 5, 9.1.4.

¹⁷⁸ Voir annexe Partie 5, 9.1.6.

¹⁷⁹ Voir annexe Partie 5, 9.2.2.

¹⁸⁰ Voir annexe Partie 5, 9.2.2.

¹⁸¹ Voir annexe Partie 5, 9.2.4.

Fernand explique que ce sont des contraintes contextuelles qui le font renoncer à poursuivre sa logique de mise en évidence de la qualité de l'apprentissage évalué jusqu'à la notation. Pourtant, il affirme aussi que la seule note qu'il peut qualifier en termes de contenu est le 4. Nous faisons donc l'hypothèse que plus ou moins implicitement, ce sont des contraintes liées aux objets de savoir qui le font s'attacher aux points, et qui lui posent donc des problèmes pour exploiter ses critères dans des échelles descriptives. Sa table de spécification, à ce titre, peut nous fournir quelques pistes explicatives.

Grille d'évaluation TS aire, périmètres & changements d'unités de longueur et de surface. MSN												
Progressions d'apprentissage	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Ex. 6	Ex. 7	Ex. 8	Ex. 9	Ex. 10	Ex. 11	Total
Comparaison, classement et mesure de grandeurs (longueur, aire, volume, angle, masse) par manipulation de lignes, angles, surfaces, ou solides, en utilisant des unités conventionnelles et non conventionnelles (1)		1					1	1	1	1	1	6
Estimation de grandeurs, choix d'une unité adéquate, prise de mesure à l'aide d'un instrument adapté et expression d'une grandeur dans diverses unités : (4, 5, 7)												0
Mesure des dimensions adéquates et calcul : (9)			1									1
du périmètre d'un polygone				1							1	2
de l'aire d'un carré, d'un rectangle, d'un triangle, d'un parallélogramme, d'un losange (par décomposition et à l'aide d'une formule)	1	1	1	1	1	1	1	C	2	1	1	12
de l'aire d'un polygone par décomposition en figures simples											1	1
tri et organisation des informations (liste, tableau, schéma, croquis,...)						1	1	a			1	4
mise en œuvre d'une démarche de résolution												0
ajustement d'essais successifs												0
pose de conjectures, puis validation ou réfutation												0
déduction d'une ou plusieurs informations nouvelles à partir de celles qui sont connues					2	1		b		0		4
réduction temporaire de la complexité d'un problème												0
utilisation des propriétés des figures et des grandeurs pour établir des preuves	1	1	1	1						0		4
vérification, puis communication d'une démarche et d'un résultat en utilisant un vocabulaire, une syntaxe ainsi que des symboles adéquats			1	1	1	1		1	1	1	1	8
												0
Précision du dessin				1								1
												0
Total	2	3	4	5	4	4	3	5	4	3	6	43

Figure 63 : Fernand, épreuve 2, géométrie, table de spécification

En effet, elle montre que la plupart des objectifs sont évalués au travers de plusieurs exercices qui sont autant de tâches complexes. Mais peut-il en être autrement ? Ou pour le dire différemment, aurait-il été possible de mettre en correspondance un exercice (ou une tâche complexe) avec un seul élément de progression des apprentissages du PER, ce qui aurait probablement facilité leur hiérarchisation, et ainsi, la construction potentielle d'une échelle progressive centrée sur les contenus ? Rien n'est moins sûr, et cela hypothétiquement pour des raisons de caractéristiques de l'objet de savoir « aires et périmètres ». Vu sa complexité et ses caractéristiques multiples, nous faisons l'hypothèse que cet objet ne peut que « rayonner » dans plusieurs tâches, ce qui serait un signe de cohérence d'un point de vue l'alignement curriculaire. Le traitement de ces multiples dimensions de l'objet de savoir, *via* plusieurs tâches évaluatives, peut donc supporter une pondération en points, mais se heurte à un traitement de ces mêmes

points pour produire des notes dans lesquelles les dimensions participent à discriminer les bons résultats des résultats satisfaisants et insuffisants.

Il nous semble que le même raisonnement peut être tenu dans la seconde épreuve de Paul portant sur « Fonctions et algèbre : calcul littéral », dans la mesure où elle est organisée selon les mêmes principes que celle de Fernand, y compris au niveau de sa table de spécification, reproduite ci-dessous.

Voici le tableau à utiliser pour l'évaluation. Lorsque l'appréciation pour une progression dans un exercice est insuffisante, on indique le chiffre 1. Si l'appréciation est suffisante, on indique le chiffre 2. Si la progression est considérée comme largement atteinte, on entre la valeur 3.

	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6	Ex 7	Ex 8	Ex 9	total
P1										
P2										
P3										
P4										
P5										
P6										

On peut donc obtenir au total 18 points au maximum sur l'ensemble de l'évaluation, ce qui représente la note de 6. Si toutes les progressions sont suffisantes, on obtient 12 points, c'est le 4.

Sur cette base j'ai créé l'échelle suivante :

note	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
points		6	7	8	9	10	12	13	15	16	18

Figure 64 : Paul, épreuve 2, calcul littéral, pondération et notation (tableau)

La différence entre les deux réside dans le fait que Paul qualifie chaque progression d'apprentissage du PER à l'aide de trois niveaux « insuffisant (1)-suffisant (2)-complètement acquis (3) » et qu'il les convertit en points. Cette procédure lui permet de construire le seuil du 4, mais pas les autres, qu'il déduira en les comparant à ce que le barème fédéral propose. Là, Paul affirme qu'il est trop subjectif pour lui de déterminer ce qui est justement suffisant ou complètement acquis pour chaque progression des apprentissages, et il reste en difficulté à argumenter de son choix d'utiliser trois niveaux et non deux, ce qui aurait hypothétiquement permis de s'approcher d'une note référée aux apprentissages, la tendance centrale étant potentiellement évitable. Mais de nouveau, nous faisons l'hypothèse que c'est également la circulation des caractéristiques de l'objet de savoir « Fonctions et algèbre : calcul littéral » dans

les différentes tâches évaluatives de son épreuve qui rendent cette notation délicate.

Philémon, quant à lui, n'est pas passé par la formulation de critères¹⁸². Cela ne l'empêche pas de noter en fonction des minima attendus dans chaque exercice pondéré en points, et surtout, de s'appuyer sur les progressions d'apprentissage du PER pour faire des feedbacks aux élèves sous forme de « toujours – parfois – jamais ». Dans ce sens, son épreuve, plus traditionnelle que celle de ses deux collègues, ne nous fournit pas d'éléments significatifs de réponse pour comprendre ce qui l'empêche d'aller vers une notation référée à l'apprentissage. Et cela, même s'il pose dorénavant le barème en amont.

Pour le français

D'une certaine manière, c'est bien l'épreuve d'Alizée et d'Anouck qui a provoqué le questionnement qui nous occupe ici, appuyé par nos données relevées dans les discussions des épreuves lors des séminaires de formation. Nous faisons ici l'hypothèse que c'est l'objet de savoir « production écrite d'un texte explicatif » qui leur a notamment permis de construire leur grille de critères progressifs permettant de qualifier les échelons de leur outil de notation qu'est l'échelle descriptive reproduite ci-dessous.

¹⁸² Voir annexe Partie 5, 9.1.5 et 9.1.6.

<p>Pour avoir la note correspondante, tous les éléments indiqués doivent être présents dans le texte !</p>	
6	<p>Doit comporter tous les éléments pour le 4, le 4.5, le 5 et le 5.5, mais, en plus :</p> <ul style="list-style-type: none"> - Le texte dans son ensemble est d'une grande cohérence. - Il est très original et ne contient aucune maladresse.
5.5	<p>Doit comporter tous les éléments pour le 4, le 4.5 et le 5 mais, en plus :</p> <ul style="list-style-type: none"> - le vocabulaire est varié, sans aucune maladresse. - Les arguments sont particulièrement intéressants et bien illustrés par des exemples. - Le contre-argument est réfuté de manière intéressante.
5	<p>Doit comporter tous les éléments pour le 4 et le 4.5, mais, en plus :</p> <ul style="list-style-type: none"> - le titre est intéressant. - Il y a des reprises utilisées correctement. - Il n'y a pas plus d'une répétition péjorant la qualité du texte. - Le vocabulaire est utilisé correctement, une légère maladresse est tolérée.
4.5	<p>Doit comporter tous les éléments pour le 4, mais, en plus :</p> <ul style="list-style-type: none"> - Présence d'une thèse, d'une thèse adverse et d'une conclusion vraisemblables et intéressantes. - Marques de subjectivité présentes. - Exemples et arguments intéressants.
4	<ul style="list-style-type: none"> - Présence d'une thèse, d'une thèse adverse et d'une conclusion vraisemblables. - Tous les éléments des consignes d'écriture sont présents, sans exception.
3.5	<ul style="list-style-type: none"> - Il manque un élément important du texte (argument, exemple, thèse, conclusion).
3	<ul style="list-style-type: none"> - Il manque des éléments des consignes d'écriture. - Le texte est trop court ou incohérent.
2.5	<ul style="list-style-type: none"> - Le texte est peu compréhensible, mal construit, pas clair, il présente de très importants défauts indiqués sur le texte.

Figure 65 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), échelle de notation

En effet, nous faisons l'hypothèse qu'il est plus aisé de définir ce que peuvent être minimalement les caractéristiques d'un texte suffisant (note 4), ces caractéristiques se rapportant au genre « texte explicatif ». Le jugement à porter sur les éléments complémentaires

à fournir pour obtenir un texte plus dense, pertinent ou complet, suit dès lors une progression que l'on pourrait qualifier de curriculaire, puisqu'elle se réfère aux objectifs évalués et aux caractéristiques du genre de texte. Certes, la démarche pourrait être qualifiée de cumulative, puisque pour obtenir un 4.5, l'élève doit maîtriser tous les éléments requis pour le 4, plus trois autres, et ainsi de suite. Toutefois, cette logique est selon nous davantage centrée sur la recherche de preuves d'apprentissage¹⁸³ à valider que sur l'addition d'éléments éloignés du contenu. Cette hypothèse se confirme dans la construction du seuil du 4, qui se réfère à des caractéristiques de l'objet de savoir que les deux enseignantes ont considéré comme incontournables.

C'est donc bien un raisonnement didactique sur l'objet qui est selon nous en jeu ici pour évaluer, et *in fine* construire la note. Et c'est hypothétiquement encore ce raisonnement didactique sur l'évaluation et sur la note qui est à nos yeux encore en jeu ici lorsque les deux enseignantes décident d'extraire de l'évaluation de la production écrite les objets de savoir liés au fonctionnement de la langue (orthographe, conjugaison, syntaxe), pour en faire une seconde note, de TA cette fois-ci. Ainsi, avec une même épreuve, Alizée et Anouck obtiendront deux notes portant sur des objets certes articulés, mais didactiquement distincts. Est-ce l'apanage des langues ? Est-ce spécifique à certains objets ? Nous reviendrons sur ces interrogations fondamentales dans notre conclusion.

<u>Technique de la langue : ETA</u>		<u>Note :</u>
On tiendra compte des 100 premiers mots pour l'ETA. ➤ N'oubliez pas d'indiquer le 100 ^e mot par un *		
Orthographe :	0 à 2 erreurs : 2 pts 5 à 6 erreurs : 1 pt 9 erreurs ou plus : 0	3 à 4 erreurs : 1.5 pt 7 à 8 erreurs : 0.5 pt
Conjugaison :	0 à 1 erreur : 2 pts	2 à 3 erreurs : 1 pt
Syntaxe (phrases correctes, connecteurs utilisés correctement, ...) et ponctuation :	0 à 1 erreur : 2 pts 3 à 4 erreurs : 1 pt 6 erreurs ou plus : 0 pt	2 erreurs : 1.5 pt 5 erreurs : 0.5 pt

¹⁸³ Nombre de chercheurs anglo-saxons que nous citons dans cette recherche, comme par exemple Wiggins et McTighe (2005) ou Guskey, Swan et Lee Ann Jung (2011), parlent d'« *evidence* » comme synonyme de preuve.

Figure 66 : Anouck et Alizée, épreuve 2, expression écrite (le texte explicatif), pondération et notation du TA de fonctionnement de la langue

Nous l'avons dit, la seconde épreuve de Samia, quant à elle, porte sur des objets de savoir peu complexes. Il eut été probablement plus simple pour elle de la considérer comme telle, et ainsi de la pondérer et de la noter de manière traditionnelle. Il nous est dès lors difficile d'en tirer des pistes de réflexion relativement à ces questions d'obstacles à une notation référée à l'apprentissage, puisqu'au travers des tâches, puis des points et de l'échelle, l'épreuve ne donne pas suffisamment de preuves d'apprentissages significatifs dans la discipline.

2.2.4 Les obstacles à la notation : qu'en est-il ?

La lecture des pages précédentes pourrait laisser penser que tous ces obstacles sont indépendants les uns des autres. Nos interprétations tendent plutôt à montrer qu'ils se conjuguent, rendant leur dépassement vraisemblablement encore plus délicat, et cela d'autant plus qu'ils se situent selon nous à deux niveaux.

Le premier renvoie à des décisions politiques, légales, réglementaires. Il concerne la loi scolaire, son règlement d'application, le cadre général de l'évaluation, voire parfois même les directives d'établissement, qui, de concert, formalisent et décident du statut de la note à l'école, de ses modes de construction et de son emploi. Là-dessus, les enseignants n'ont aucune prise. Ou presque. La preuve : nous observons que certains d'entre eux prennent avec ces nombreuses prescriptions une distance bienvenue et assumée, gage de professionnalisation, pour autant que leur contexte d'établissement le leur permette. En effet, nous faisons l'hypothèse que réduire l'empan des échelles de notation par exemple - ce qui est, formellement, illégal - ne peut se réaliser que dans un établissement où le leadership est peu contrôlant, et où les équipes sont relativement autonomes, comme cela semble être le cas pour les enseignantes de français. Ce niveau appellerait donc selon nous des choix essentiellement de l'ordre du compromis.

Le second niveau est celui des pratiques possibles. Là, le rôle des connaissances clés pour progresser vers une notation référée à des contenus est souligné par nos résultats. On y observe que les enseignants, pour aller dans ce sens, mobilisent notamment et non seulement des apports sur les objets de savoir, sur le plan d'études, sur les tâches, sur les systèmes de pondération, mais surtout, et c'est ce qui est à nos yeux marquant, ils exploitent des raisonnements sur les liens que ces différentes composantes entretiennent entre elles. Notre hypothèse est que ce raisonnement systémique est possible car il repose sur une référence au modèle de l'alignement

curriculaire élargi. Cela ne signifie en rien une application du modèle, fort heureusement puisque ce n'est pas notre visée, mais rend la compréhension des phénomènes de cohérence en évaluation sommative notée plus appréhendable pour les praticiens. Ce qui nous amène à traiter de notre troisième sous-question de recherche, puisque nous postulons que les connaissances acquises par les enseignants, puis leur exploitation en contexte sous forme de compétences, sont possibles grâce au modèle théorique de l'alignement curriculaire élargi.

2.3 Trois apports sur les pratiques de notation du point de vue de l'alignement curriculaire élargi

L'étude et la compréhension des pratiques d'évaluation sommative notée saisies dans notre modèle fournit selon nous trois apports clés.

- Apport 1 : penser les pratiques de notation au cœur du modèle de l'alignement curriculaire élargi permet aux enseignants de penser de manière articulée d'autres dimensions de leur pratique évaluative, voire des dimensions relatives aux processus d'enseignement et d'apprentissage.

Lorsque les enseignants pensent les seuils de suffisance en fonction des contenus clés à maîtriser, quand ils analysent la pertinence du nombre d'échelons, ou lorsqu'ils décident de poser leurs barèmes en amont de l'épreuve et chaque fois de manière spécifique à la situation d'enseignement-apprentissage et d'évaluation, ils mobilisent des connaissances qui n'ont pas, directement, de lien avec la notation. Ainsi, par exemple, et sans vouloir revenir sur des données déjà amplement discutées, nous voyons qu'au travers de la recherche de ce que pourrait être le contenu à valider pour décider du seuil de la note 4, un raisonnement sur l'objet de savoir est souvent en jeu, et parfois même déjà en amont de la situation d'évaluation sommative. C'est ce même raisonnement qui va infléchir les processus d'enseignement et d'apprentissage, et donc, qui va hiérarchiser le sens des tâches d'apprentissage puis d'évaluation, puis corollairement les décisions qui vont décider du poids de l'apprentissage évalué dans les épreuves. Dans ce cas de figure, la notation vient formaliser la transaction entre ce complexe raisonnement et l'attribution d'un chiffre, à la différence près que ce chiffre n'est plus déterminé par un outil extérieur aux apprentissages, mais par un outil créé idéalement en continuité avec le contenu de l'apprentissage enseigné, appris, puis évalué. La dimension systémique de notre modèle théorique « élargi » semble ici bien se confirmer.

- Apport 2 : penser les pratiques de notation au cœur du modèle de l'alignement curriculaire élargi permet aux enseignants de penser dialectiquement évaluation et notation.

À la lumière de nos précédents résultats, nous confirmons que le processus de notation entretient bel et bien un fort lien de cohérence avec le processus d'évaluation sommative - lui-même en lien avec les processus d'enseignement et d'apprentissage. Ce constat nous semble particulièrement pertinent d'un point de vue de recherche et de formation, pour trois raisons au moins :

- Il donne à la pratique de notation un statut fondamental, complexe, aux antipodes d'une pratique mécaniste, voire technique, se résumant à l'exploitation d'outils standardisés ou dénués de rapport avec les contenus enseignés et appris ;
- Dans ce sens, ce constat met bel et bien en évidence qu'il est possible, pour un enseignant, d'être aligné curriculairement entre la référenciation restreinte et la pondération, puis potentiellement désaligné lorsqu'il s'agit de noter. D'un point de vue conceptuel, cet apport rendu possible par notre modèle est particulièrement intéressant, car il met en évidence la complexité de son acception « élargie », mais aussi sa difficulté à être exploité dans certains contextes.
- Il devient primordial de revoir les perspectives « *bottom-up* » ou « *top-down* » du modèle de l'alignement curriculaire également dans sa conceptualisation élargie. Dans ce sens, nous serions tenté de dire qu'à l'instar des six enseignants de notre recherche, un praticien devrait pouvoir tout autant questionner une cohérence curriculaire en entrant par la note dans une perspective ascendante (dans quelle mesure la note construite ou à construire renvoie-t-elle à des unités d'apprentissage ?), que d'interroger cette cohérence depuis le curriculum jusqu'à la note dans une vue descendante (à quelles conditions les caractéristiques des objets de savoir et des objectifs pourront circuler dans les tâches, les systèmes de pondération et la construction de la note ?). Si la seconde est intelligible, la première, à notre sens, suppose un nouveau questionnement professionnel et de recherche, comme celui que nous proposons ici. Ce questionnement depuis la note devrait tout aussi bien être au service d'une cohérence pédagogique que l'analyse du curriculum ou de l'objet de savoir, ce qui, à notre connaissance, est également une façon relativement innovante de penser la question.

- Apport 3 : le contexte de la pratique évaluative a une forte influence sur la viabilité du modèle « élargi » comme outil compréhensif.

Une perte de cohérence est donc possible entre évaluation et notation, pour des raisons pédagogiques, didactique et/ou évaluatives. Mais aussi pour des raisons contextuelles. Ainsi, voir des enseignants renoncer à cette recherche de cohérence plus ou moins volontairement au moment de noter et se rabattre sur des procédures standardisées, faute de mieux, donne à penser que le contexte joue ici un rôle déterminant dans la (non) évolution des pratiques de notation. Mais conjointement – et c’est selon nous un apport également majeur de cette recherche – nous constatons la puissance de questionnement du contexte qu’offre le modèle. En effet, la recherche de cohérence curriculaire qui en est le centre ne cesse d’entrer en confrontation avec la nature des outils proposés ou imposés de l’extérieur (les barèmes standardisés et leur linéarité, les moyennes cible dans certains établissements, l’élaboration des seuils en pourcentages, pour ne citer que ces exemples). Ainsi, à défaut de donner parfois des réponses pour la pratique, le modèle permet une meilleure appréhension et compréhension des outils de notation et de leurs directives institutionnelles.

3 Conclusion

Ce chapitre confirme que les pratiques de notation préoccupent les enseignants, sont loin d’être mécanistes et relèvent de processus de jugements très complexes. Nous avons pu mettre en évidence les tensions, voire les dilemmes que rencontrent les praticiens quand ils font état de leurs pratiques de notation spontanées, avant une formation. Et ces pratiques sont donc loin d’être manichéennes : les propos disant que les enseignants aiment la note parce qu’elle les rassurerait, leur donnerait du pouvoir ou leur permettrait de rendre des comptes, ou qu’ils l’abhorrent parce qu’elle serait un outil au service d’une forme d’asservissement des élèves et de leur démotivation doivent donc être nuancés, surtout lorsque l’on se place du point de vue de l’étude des pratiques réelles.

En effet, les pratiques de notation des enseignants de cette recherche sont bien plus complexes que cela, bien moins contrastées. À la lumière de nos résultats, nous sommes tenté de dire qu’elles traduisent vraisemblablement une recherche de compromis entre une volonté ou un devoir de responsabilité ou « de rendre compte », et un besoin de fournir des informations porteuses de sens pour le parcours de l’élève.

La formation construite et dispensée autour du modèle théorique de l'alignement curriculaire élargi semble leur permettre d'assumer relativement bien ces compromis : elle améliore les pratiques de notation, mais ne résout pas tous les dilemmes ni les tensions. Elle semble davantage participer à les identifier et à les gérer, ce qui est selon nous un premier résultat important. Dans ce sens, n'est-il pas interpellant de voir des enseignants noter en étant dépourvus d'outils pour élaborer leur propre processus de construction de notes ? Peut-on continuer à défendre l'idée que l'enseignant est compétent pour enseigner, gérer l'apprentissage, évaluer, et lorsqu'il s'agit de noter, accepter qu'il abandonne un jugement situé et remette ses choix dans les mains d'outils qui n'ont aucun rapport avec l'immense travail qu'il assume au quotidien avec ses élèves ? C'est pourtant bien ce qui se passe dans nombre de classes où les pratiques de notation répondent davantage aux directives des décideurs et des chefs d'établissements qu'à des soucis de cohérence pédagogique. Car nous l'avons mis en évidence dans ce chapitre : ce sont bien ces mêmes directives et croyances qui semblent freiner le développement de pratiques de notation constructive chez certains participants à la recherche.

Un questionnement plus global, suscité par les malaises des enseignants à noter et par ces premières réflexions, peut alors selon nous voir le jour. Il renvoie à un paradoxe plus général et hypothétiquement lié au contexte des enseignants, qui interroge au moins deux axes :

- Comment se fait-il que l'évaluation sommative notée, dont une des caractéristiques est de posséder une dimension sociale, soit l'objet de tant de prescriptions de la part des décideurs et des praticiens ? Ne serait-ce justement pas là que la particularité des situations d'évaluation devrait être aussi perceptible, avec l'usage d'outils chaque fois spécifiques ?
- Conjointement, comment expliquer que les enseignants soient si insuffisamment formés à cet exercice complexe, puisqu'il en va de la pertinence des décisions qu'ils prennent pour les parcours des élèves, et plus largement, de leur développement professionnel ?

Ce questionnement est d'autant plus important qu'il met en évidence l'impact positif que la formation, à certaines conditions et à l'instar de ce que nous avons mené dans cette recherche, peut avoir sur les pratiques. Nous y reviendrons dans notre partie conclusive. Toujours est-il qu'il est possible d'améliorer les pratiques de notation, ou en tous cas de permettre leur conscientisation par les enseignants.

Un travail à réaliser pour appréhender ces paradoxes et impossibilités revient pour nous à penser comment contourner ou aborder les trois obstacles au développement de pratiques de notation

cohérentes que nous avons mis en évidence, à savoir :

- Relativement à l'obstacle du barème et des échelles : penser que les faiblesses des échelles n'existent que lorsqu'elles sont formulées sur la base de totaux de points. En effet, lorsque l'échelle devient descriptive, elle constitue un outil qui rend compte de l'apprentissage évalué. À ce titre, elle n'obéit plus à des impératifs de la mesure qui sont à l'origine des désalignements curriculaires qu'elle peut induire, comme nous le montrons dans notre cadre conceptuel et dans notre précédent chapitre de résultats.
- Relativement à l'obstacle des seuils, et notamment à celui du seuil de réussite : problématiser le fait que c'est ce dernier qui reste prépondérant dans ce travail d'élaboration complexe, et qu'à ce titre, nous faisons l'hypothèse que moins il y a de seuils à fixer, plus ils seront aisés à déterminer.
- Relativement à l'obstacle de l'objet de savoir évalué : concevoir qu'hypothétiquement, il existe des objets de savoir dont les caractéristiques, d'un point de vue didactique et évaluatif, sont davantage systémiques, alors que d'autres, semblent être plutôt « cumulatifs ». Ces différents modes de structuration infléchissent les possibles dans les pratiques de pondération et de notation.

D'un point de vue général donc, et fidèlement à notre modèle théorique et à nos résultats, nous souhaitons ici affirmer que le processus de notation doit être considéré comme étant étroitement articulé au processus d'évaluation, tout en s'en distinguant. Pour étayer encore cette affirmation, nous nous appuyons sur l'observation des pratiques qui parviennent à questionner les processus d'enseignement, d'apprentissage, ainsi que les démarches de référenciation restreinte, de *design* et de pondération à partir des échelles descriptives, ce qui, au début de notre recherche, n'était qu'une hypothèse. À notre sens, cet apport ouvre des possibles conséquents en termes de recherche (en problématisant encore davantage les liens entre évaluation et notation selon les contextes et les objets de savoir évalués) et de formation (en articulant des modèles théoriques avec les pratiques et questions réelles des enseignants).

Toutefois, un problème de taille perdure dans de telles démarches, celui du statut du curriculum de référence à partir duquel l'enseignement, l'apprentissage et l'évaluation sont censés être développés. En effet, comment penser la cohérence de l'évaluation, et de surcroît de la notation, lorsque l'absence d'un référentiel d'évaluation se fait sentir ? Ne nous trouvons finalement pas ici face à l'obstacle majeur d'une évaluation et d'une notation alignées curriculairement ? Nous en débattons dans notre conclusion.

Partie 9 - Conclusion

1 Propos liminaires

Dans cette recherche doctorale, nous avons investigué une problématique déclinée au travers de trois dimensions dans nos questions de recherche : les points sur lesquels les enseignants portent leur attention lorsque leurs pratiques évaluatives sommatives sont pensées à l'aune du modèle théorique de l'alignement curriculaire élargi, les compétences évaluatives que ce modèle participe à développer, et l'investigation des pratiques de notation saisies dans ce modèle.

Nous commencerons donc cette conclusion par une synthèse de nos principaux résultats, que nous tenterons le plus possible de mettre en dialogue à l'aide de références issues de la recherche sur lesquelles nous nous sommes appuyé, mais aussi avec de nouvelles, complémentaires.

La deuxième partie portera sur trois discussions majeures, et selon nous nécessaires, émanant de cette recherche :

- Une discussion conceptuelle visant à mettre en évidence l'évolution de notre modèle théorique une fois soumis aux pratiques réelles et aux différentes dimensions des contextes dans lesquels ces dernières s'inscrivent, ainsi que l'impact du modèle sur les pratiques évaluatives ;
- Une discussion méthodologique sur les atouts et biais des méthodes choisies ;
- Corollairement, une discussion autour de l'articulation des rôles de formateur et de chercheur que nous avons assumés, et de notre possible posture d'accompagnateur telle que nous l'avons évoquée dans notre ancrage méthodologique (partie 6).

Suivra une réflexion faisant état de certaines limites de notre recherche dans une perspective critique et réflexive. Nous poursuivrons en proposant quelques pistes possibles suite à notre travail, du point de vue de la recherche et de la formation. Enfin, tenter de développer des compétences évaluatives d'enseignants en exercice au travers d'un dispositif comme le nôtre pose la question de savoir dans quelle mesure ce travail comprend, *in fine*, un enjeu majeur de développement professionnel. Nous concluons sur ce point.

2 Les principaux apports de notre recherche

Nous recensons ces apports en revenant à nos trois questions générales de recherche.

2.1 Les dimensions des pratiques évaluatives sur lesquelles les enseignants portent leur attention lorsqu'elles sont confrontées aux composantes du modèle théorique de l'alignement curriculaire élargi

Les portraits des enseignants mettent en évidence le fait que ces derniers sont principalement préoccupés par des questions relatives à la notation avant d'entrer en formation. Ce questionnement traduit un désir de formuler un bilan des apprentissages « juste » au regard de ce qu'ils observent sur les épreuves de leurs élèves, ce que nous assimilons à une forme de préoccupation éthique dans la mesure où l'enseignant interroge ses pratiques à l'aune de valeurs comme la justice, l'équité ou l'égalité (Jeffrey, 2013). Toutefois, si ce désir est compréhensible pour des professionnels désireux de devenir de « meilleurs » évaluateurs, le terme « juste » demande à être explicité, car il peut potentiellement recouvrir une multitude de significations en fonction des contextes dans lesquels les pratiques évaluatives s'inscrivent (Tierney, 2013). Il serait toutefois abusif d'affirmer que les enseignants ne se préoccupent que de la notation. Nous avons montré que le rapport au PER, notamment dans son statut de référentiel d'évaluation (Mottier Lopez, 2016 ; Mottier Lopez & Dechamboux, 2017), ou que le *design* possible relativement à certains objets de savoir interpellent également. Notons tout de même que si les interrogations relatives à la notation viennent spontanément, ces derniers axes sont questionnés dans un deuxième, voire un troisième temps.

Cela étant, nous avons établi que les enseignants des deux disciplines, une fois outillés par le modèle théorique, sont souvent à même de penser et de comprendre la plupart des phénomènes de cohérence au sein de leurs pratiques évaluatives. Nous avons pu montrer que, de manière générale, les participants sont en mesure de faire des choix au sein du curriculum, de concevoir le *design* de leurs tâches évaluatives en homogénéisant leur niveau de complexité du point de vue des habiletés cognitives et des contenus (Guskey, 2006), et d'élaborer des systèmes de pondération davantage en adéquation avec cette complexité (Smith, Smith & De Lisi, 2001). Ces résultats vont nous amener à revenir plus loin sur le statut du modèle théorique dans notre dispositif de recherche, et également sur le rôle d'accompagnateur que nous avons investi (Charlier, 2012). Toutefois, des objets de débat restent en suspens.

Le premier d'entre eux, et le plus interpellant à nos yeux, est le fait que l'articulation entre pratiques réelles et modèle théorique se heurte à l'impossibilité de dépasser des ruptures d'alignement curriculaire élargi au niveau de la notation. Biggs (2003) a bien mis cet écueil en évidence dans ses travaux, en montrant notamment ici le rôle de désalignement - ou de « *misalignment* » (Airasian & Miranda, 2002, p. 254) - que peut jouer l'échelle de notation standardisée. Le fait qu'Anouck et Alizée, en français, parviennent à développer des pratiques évaluatives cohérentes du curriculum à la notation ne nous permet toutefois pas d'affirmer que le modèle offre une lecture compréhensive de la cohérence évaluative dans toutes les disciplines et/ou avec tous les objets de savoir. Au contraire. Des ruptures semblent être indépassables pour différentes raisons sur lesquelles nous reviendrons.

Cela étant, le modèle théorique exerce une influence sur la manière avec laquelle les enseignants identifient et/ou évitent les ruptures d'alignement curriculaire au sein de leurs épreuves sommatives. Nous observons effectivement que leur cohérence, de manière quasi univoque, s'améliore tout au long du processus entre la référencement restreinte et le *design*, entre le *design* et la pondération, et parfois, ou de manière partielle, entre la pondération et la notation, ce qui va dans le sens de plusieurs travaux ayant thématiqué cette question (Bateman et al., 2009 ; Guskey, Swan, & Lee, Ann Jung, 2011 ; Klenowski & Wyatt-Smith, 2014). Mais il est encore plus intéressant de constater que cette cohérence s'accroît également entre la référencement restreinte et la pondération, ou entre le *design* et la notation, ce qui tendrait à confirmer que le modèle ne renvoie effectivement pas à un raisonnement linéaire, mais bien systémique. Cette nouvelle acception renforce l'inscription du modèle « élargi » dans deux composantes du modèle de base : les processus d'enseignement et d'apprentissage. Ce constat asseoit selon nous le fait que l'évaluation sommative notée est partie intégrante de ces mêmes processus d'enseignement et d'apprentissage (Brookhart, 2017). D'un point de vue conceptuel, nous serions donc tenté de dire que l'acception « élargie » du modèle est théoriquement pertinente. Ce constat s'est confirmé dans nos données lorsque, par exemple, des participants comme Anouck, Samia ou Paul nous ont dit ne plus enseigner de la même manière suite à la formation. Ce sont donc bien ces mêmes données relevant de pratiques réelles qui ont questionné le modèle, en ont montré les atouts et aussi les limites que nous débattons plus bas.

2.2 Les compétences évaluatives spécifiques d'un point de vue sommatif que l'alignement curriculaire élargi participe à développer

Les résultats obtenus ici sont selon nous intéressants et viennent enrichir les précédents. Nous

observons que les compétences qui se développent de manière significative chez tous les praticiens sont les deux premières, à savoir :

- A. Analyser le plan d'études du point de vue des objectifs et des objets de savoir disciplinaires qui y circulent à des fins d'évaluation sommative en mettant en relation de manière cohérente les différents niveaux de formalisation des objectifs, et en analysant les objets de savoir pour en déterminer les caractéristiques essentielles à des fins d'évaluation sommative.
- B. Elaborer des tâches évaluatives complexes en rapport avec les objectifs des plans d'études qui donnent des preuves d'apprentissages maîtrisés ou en voie de réalisation, en privilégiant des tâches qui donnent à voir des unités d'apprentissage significatives (approche qualitative).

Nous constatons donc que le modèle, initialement conceptualisé pour penser et comprendre les pratiques évaluatives sommatives, participerait à développer une compétence d'analyse du curriculum et une autre de construction de tâches évaluatives. Ce résultat permet deux réflexions au moins :

- Les compétences évaluatives sommatives se développent conjointement à d'autres compétences professionnelles (Edwards, 2017). En effet, développer ses pratiques évaluatives sommatives revient pour le praticien à mobiliser des connaissances qui relèvent de dimensions présentes en amont de cette pratique ;
- Corollairement, penser ses pratiques évaluatives sommatives notées oblige l'enseignant à revisiter le système dans lequel elles s'inscrivent (Brookhart, 2017). Il s'agit donc pour lui de réfléchir aux liens de cohérence qui devraient exister entre les processus d'enseignement, les processus d'apprentissage et les tâches y relatives, les prises d'information à visées formatives, puis les démarches d'évaluation sommatives. Nous retrouvons ici les éléments définitoires de l'alignement curriculaire d'Anderson (2002).

Pour nous, cette articulation entre les différentes dimensions de la pratique évaluative résonne avec la perspective systémique de notre modèle théorique : à partir du moment où les questions des enseignants sont soumises à la rigueur réflexive que ce dernier requiert, on observe que les connaissances acquises s'inter-relient. En cela, elles participent au développement de compétences larges (Brookhart, 2011).

2.3 L'évolution des pratiques de notation chez les enseignants

Sans vouloir réduire la complexité de cette question, nous sommes tenté de dire qu'il y aurait deux pratiques de notation en vigueur chez les participants à la recherche à la fin du processus de formation, et qu'elles entretiennent avec la logique du modèle théorique de l'alignement curriculaire élargi des rapports contrastés.

La première consiste à considérer la notation, *in fine*, d'un point de vue quantitatif. Certains enseignants, comme Paul ou Fernand, conservent cette logique malgré la formation, lorsqu'ils finissent par établir des barèmes et construire des échelles à l'aide d'outils standardisés. Nous avons vu que bien souvent, leurs raisonnements, au départ, sont en partie qualitatifs dans la mesure où les notes (surtout le 4) sont pensées du point de vue de l'objet de savoir ou des objectifs, mais que leurs pratiques de notation, notamment au niveau des outils qu'ils construisent, finissent par aplanir ce raisonnement en produisant d'autres notes partiellement désalignées. Ces résultats montrent notamment que même si les enseignants connaissent les principes d'une notation cohérente, ils peinent parfois à les mettre en œuvre (Randall & Engelhard, 2010), ou que la compensation des points ou des critères, malgré son absurdité pédagogique (Marzano & Heflebower, 2011), peut être assumée par le praticien pour régler certains dilemmes. Une frustration relative à l'impossibilité de dépasser des obstacles relatifs à la pratique de notation (Linn & Miller, 2005) serait donc potentiellement bien présente ici chez certains enseignants comme Anouck, Alizée ou encore Fernand, qui l'expriment clairement dans leur dernier entretien et parfois dans leurs textes.

La seconde pratique, présente en français et sur des objets de savoir circonscrits, montre que la note, à certaines conditions, peut être référée à l'apprentissage au travers d'échelles descriptives (Laveault & Miles, 2008 ; Mottier Lopez & Laveault, 2008). Nous avons exemplifié que dans cette situation, les désalignements curriculaires peuvent, à certaines conditions également, être évités, mais que la généralisabilité de telles pratiques est délicate. Ces éléments contrastés nous ont permis de formuler quatre apports significatifs :

- Les pratiques de notation, à partir du moment où elles se pensent du point de vue de l'apprentissage évalué, renvoient davantage à un processus complexe qu'à l'aboutissement d'une démarche mécaniste. Certes, ce résultat est vraisemblablement un effet de notre dispositif. Mais il vient tout de même corroborer nombre de constats issus de recherches sur le « *grading* » (e. g., Brookhart, 2013 ; Guskey, 2006, 2011 ; Walvoord & Johnson Anderson, 2009) ;

- Ce processus complexe ne peut se déployer dans une recherche de cohérence d'alignement curriculaire élargi qu'à la condition où les enseignants sont en mesure de construire leurs propres échelles de notation en référence à l'enseignement dispensé, aux opportunités d'apprentissage offertes aux élèves, à la complexité des objectifs et des objets de savoir, et à leur contexte (Biggs & Tang, 2009, 2011 ; Brookhart, 2017). Afin de réaliser ce travail d'une complexité conséquente, ils devraient connaître les caractéristiques des objets de savoir incontournables et à maîtriser par les élèves (Moss, 2013). Par ailleurs, il semble se confirmer que le problème à construire une note référée aux apprentissages est bien le nombre d'échelons imposés aux enseignants (Marzano & Heflebower, 2011 ; Reeves, 2004), comme nous l'avons évoqué dans notre partie théorique sur les pratiques évaluatives¹⁸⁴. Le problème est que les travaux dont nous disposons sur cette question, comme ceux de Biggs (2003) ou de Royce Sadler (2009), s'inscrivent dans des contextes où les échelles sont toujours identiques (notes A à F, avec une seule note insuffisante, F). Or, c'est bien lorsqu'elles changent d'un contexte à l'autre que le raisonnement sur la cohérence de l'alignement curriculaire élargi rencontre des obstacles. Ces travaux n'offrent donc pas de conceptualisation suffisamment développée pour penser l'évaluation sommative notée conditionnée par différents outils, raison pour laquelle nous avons opté pour une conceptualisation élargie du modèle - pertinente si l'on en croit nos résultats - pour penser cette cohérence jusque dans la notation ;
- Les caractéristiques de l'objet de savoir entrent donc également en jeu dans cette possibilité, ou non, de construire une note référée à l'apprentissage. Certains de ces objets, didactiquement plus « cumulatifs », semblent ouvrir davantage de possibles pour construire des échelles descriptives progressives, ce que montrent par exemple les travaux de Marzano et Heflebower (2011) avec les « *proficiency scales* » (p. 37) ;
- Malgré ces réserves, le modèle théorique de l'alignement curriculaire élargi est selon nous pertinent, même s'il ne peut faire fi de l'idéologie de la mesure présente dans les directives légales encourageant fortement les enseignants à utiliser des barèmes extérieurs aux apprentissages des élèves et emprunts de caractéristiques a-pédagogiques comme la linéarité (Dubus, 2006 ; Reeves, 2004). En effet, le modèle, à certaines conditions que nous avons exposées, outille les praticiens pour élaborer des notes cohérentes. Et lorsque des contraintes ne le permettent pas, il leur offre tout de même la

¹⁸⁴ A ce titre, il serait intéressant de mener une recherche du même type en France, où les notes vont de 1 à 20.

possibilité de questionner la pertinence des outils prescrits et de trouver des voies de dépassement des normes imposées. C'est par exemple le cas lorsque les six participants, dès le deuxième séminaire de formation, établissent le seuil du 4, donc de la suffisance, non pas en fonction d'un pourcentage de points arbitraire, mais en lien avec le contenu. Cette note, dont l'importance est majeure pour l'élève, pourrait dès lors être considérée comme une « *borderline grade* » (Randall & Engelhard, 2010, p. 1378), dans la mesure où sa construction attire particulièrement l'attention des enseignants. C'est donc pour nous un résultat significatif et prometteur, et cela même si les autres notes sont ensuite élaborées à l'aide d'outils extérieurs aux apprentissages. Dans ce sens, c'est bien une capacité à réfléchir et à penser sa pratique évaluative qui se développe ici, quand bien même cette pratique est parfois bridée par des normes peu en lien avec l'évaluation des apprentissages des élèves.

La conceptualisation élargie de notre modèle théorique relève la complexité des pratiques des notation et permet aux enseignants d'opérer des jugements évaluatifs faisant référence aux objets de savoir jusque dans la notation, ce qui est nouveau. Toutefois, nous observons qu'une notation holistique (*holistic grading*) (Biberman-Shalev et al., 2011 ; Royce Sadler, 2009) telle que nous l'avons conceptualisée dans notre partie 3 n'est jamais mise en oeuvre. Même si les jugements des enseignants pour construire les notes sont réalisés au regard des composantes du modèle et de leurs rapports de cohérence, les échelles qu'ils élaborent, même descriptives, restent majoritairement d'orientation analytique. Notre hypothèse de compréhension de ce constat est que la démarche holistique relève d'habitudes ancrées dans des cultures évaluatives spécifiques, notamment en ce qui concerne les échelles de notation. Par exemple, il existe de grandes différences entre « nos » échelles et celles en vigueur dans les pays d'où proviennent les recherches sur le *grading*, comme nous l'avons dit dans notre cadre conceptuel (partie 3).

Grâce à nos trois questions de recherche, nous observons que les enseignants, progressivement, étayent et justifient petit à petit leurs choix en matière d'évaluation en mobilisant plus ou moins explicitement le modèle théorique, ce qui leur permet d'avoir une assise dans leurs décisions, jusque dans la notation. Leurs jugements évaluatifs ne sont plus de l'ordre de l'intuitif ou relatifs à la prescription ou aux habitudes, mais relèvent d'un raisonnement pédagogique, ce qui selon nous est un signe de développement professionnel. Toutefois, nous constatons également que certains apports théoriques sont malmenés par nos données.

2.4 Pour ouvrir la discussion

Au vu de ce qui précède, une discussion de notre modèle s'avère nécessaire. Nous poursuivrons par une discussion d'ordre méthodologique, non sans rapport également avec le modèle qui a en partie influencé nos choix à ce niveau. Puis nous terminerons en thématissant les rôles de formateur et de chercheur et leur articulation complexe au sein d'une posture d'accompagnateur, dans une perspective réflexive, propre à toute recherche interprétative.

3 Discussion des résultats

3.1 Un modèle théorique en mouvement

Dans notre cadre conceptuel, nous nous sommes distancé d'une acception que nous avons comprise comme trop linéaire du modèle de l'alignement curriculaire (Anderson, 2002 ; Gauthier et al., 2005). Nous en avons donc proposé une acception élargie, systémique, que nous avons confrontée aux pratiques réelles.

Nos résultats montrent que cette perspective systémique, même si elle se confirme, doit encore être discutée et affinée¹⁸⁵. En effet, il demeure que certains liens entre les composantes du modèle sont toujours malmenés par les pratiques réelles et contextualisées des enseignants. L'exemple le plus significatif pour nous se situe dans les pratiques de pondération. Par exemple, lorsque les enseignants ont été amenés à critérier leurs épreuves, la majorité d'entre eux n'a pas réalisé ce travail en référence aux *design* (tâches évaluatives), mais bien par rapport aux objectifs. Cette pratique, au demeurant cohérente, va dans le sens de bon nombre de travaux sur l'évaluation critériée (e. g., Jonsson, 2014 ; Moskal & Leydens, 2000 ; Mottier Lopez, 2016 ; Reddy & Andrade, 2010), et incarne véritablement à nos yeux cette nouvelle logique systémique au cœur du modèle. Aussi en proposons-nous une schématisation revue.

¹⁸⁵ Ce qui confirme les réserves que nous avons évoquées au point 5 de notre revue de littérature sur la première modélisation de l'alignement curriculaire élargi (Partie 2).

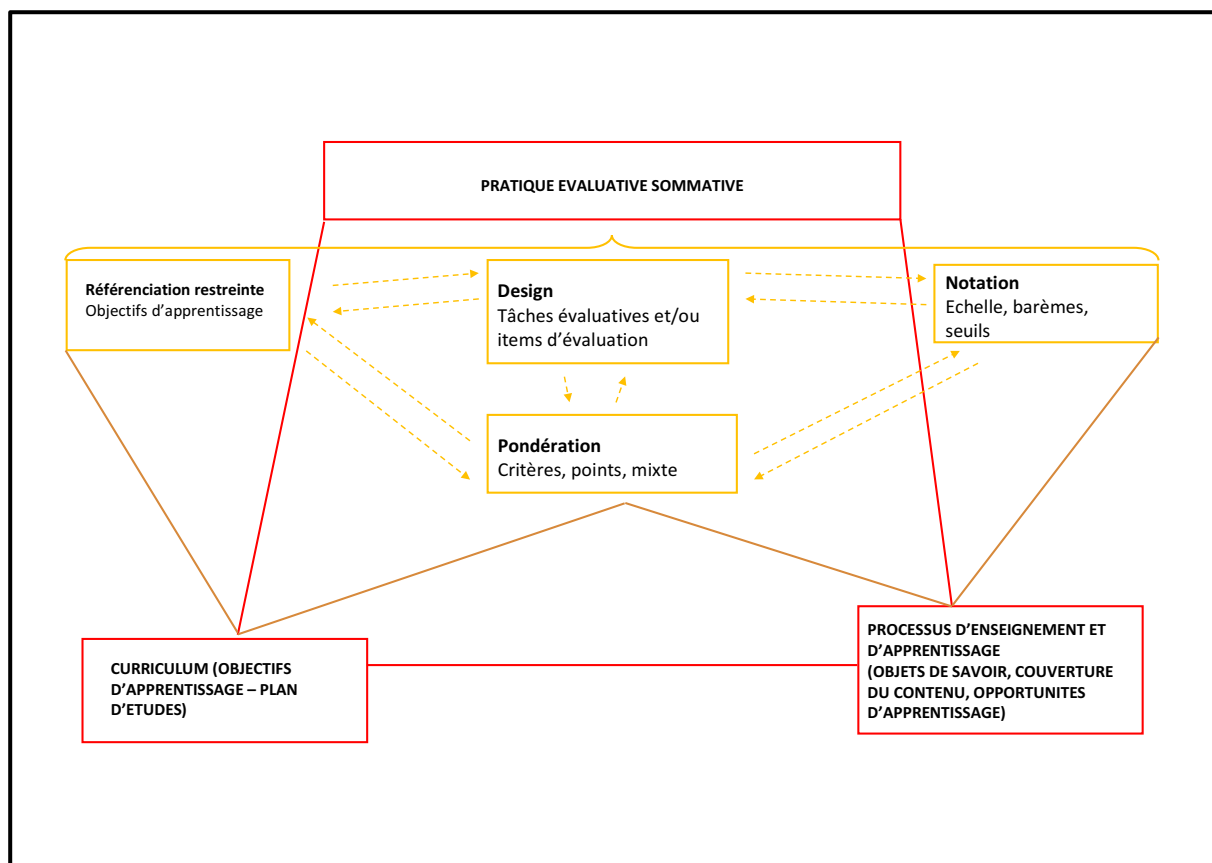


Figure 67 : Modélisation systémique de l'alignement curriculaire élargi

Nous gardons ici l'inscription du modèle « élargi » au cœur d'un modèle premier triangulaire où interagissent les processus d'enseignement et d'apprentissage, les objectifs prescrits et les démarches d'évaluation (traits rouges) (Anderson, 2002). Le modèle « élargi » déploie ensuite les quatre composantes relatives à la pratique évaluative sommative en les mettant en lien (traits jaunes). On observe alors que les rapports qu'entretiennent ses différentes composantes traduisent plus fidèlement les pratiques réelles et la rigueur conceptuelle du modèle de départ. Par exemple, nous voyons que la pratique de pondération peut se penser au regard directement des objectifs évalués dans une perspective de notation, tout en entretenant un rapport de cohérence mais non linéaire avec les tâches évaluatives (Brookhart, 2005 ; Smith, Smith & De Lisi, 2001). De manière complémentaire, on observe que la notation référée à l'apprentissage dépend non seulement de la nature des critères, mais aussi du *design* (Brookhart, 2017), puisque comme nous l'avons montré et à l'instar des préoccupations des enseignants, des tâches peu référées aux objets de savoir traduisent un alignement pauvre - « *a poor alignment* » (Biggs, 2011, p. 163) - qui ne permettra vraisemblablement en rien de noter en référence à ces mêmes objets. Dans cette nouvelle acception systémique, *la cohérence curriculaire est donc toujours minimalement le fruit d'une triangulation entre trois composantes du modèle*, ce qui est

conceptuellement nouveau.

Certains résultats plus fins viennent encore enrichir cette nouvelle perspective systémique.

Nous pensons notamment aux points suivants :

- Nous avons vu que la plupart des ruptures d'alignement curriculaire « élargi » se situent au niveau de la construction de la note lorsque cette dernière est réalisée à l'aide de barèmes extérieurs aux apprentissages. Ce résultat peut ici toujours être lu du point de vue des ruptures entre pondération et notation, mais avec une pondération cette fois-ci référée aux objectifs et chaque fois singulière, ce qui remettrait notamment en question les différentes catégories de critères que nous avons présentées dans notre partie conceptuelle¹⁸⁶. Cette réflexion complète par ailleurs selon nous les apports de Biggs (2003) dans la mesure où elle conforte le fort potentiel de désalignement que possèdent des outils comme les barèmes standardisés : elle permet de mieux comprendre les causes de ces désalignements au regard des objectifs d'apprentissage évalués ;
- Du point de vue des enseignants, le recours au modèle a un impact sur des dimensions de la pratique qui ne sont pas directement liées à l'évaluation sommative : compréhension plus approfondie du curriculum, *design* des tâches évaluatives plus homogène (Guskey, 2006), pondération référée aux contenus, remise en question de certains processus d'enseignement, ce qui validerait notre nouvelle modélisation systémique ;
- Corollairement, nous comprenons mieux les raisons pour lesquelles les compétences évaluatives qui semblent le plus se développer sont celles qui traitent de la référencement restreinte et du *design* ;
- Toutefois, il y a une impossibilité, parfois, à construire une note référée à l'apprentissage, et cela même si la cohérence d'alignement curriculaire élargi est perceptible jusqu'à la pondération. Ce problème est dû également à des raisons contextuelles et à des caractéristiques de l'objet de savoir.

Ces divers éléments, longuement débattus, nous permettent d'encore mieux appréhender la complexité des pratiques évaluatives sommatives et, conjointement, de mieux saisir la complexité du modèle « élargi ». En cela, ils devraient être considérés comme des points d'attention pour de futures recherches, voire pour la conception d'actions de formation. Il s'agit

¹⁸⁶ En effet, nous faisons l'hypothèse que la complexité des pratiques va à l'encontre de la catégorisation des critères que nous avons retenue : *task-specific rubrics* (Brookhart, 2005), *analytic rubrics* (Jonsson & Svingby, 2007), *holistic rubrics* (Isenhour & Kramlich, 2008) et *annotated holistic rubrics* (Meier et al., 2006).

toutefois maintenant de les penser en fonction du contexte dans lequel le modèle s'inscrit.

A ce titre, et comme nous l'avons déjà évoqué, le contexte vaudois qui est le nôtre est très prescriptif en matière d'évaluation sommative notée. Dès lors, que devient notre modèle théorique dans un tel environnement scolaire ? Quelles sont ses conditions de cohabitation avec les curricula, les référentiels et les outils qu'utilisent les enseignants au quotidien, notamment pour noter ?

3.1.1 Le référentiel d'évaluation au sein du modèle théorique

Nos résultats montrent que si les participants se réfèrent le plus souvent au PER pour leur enseignement, le pas à franchir pour l'exploiter dans l'évaluation sommative reste délicat. Nous pourrions voir dans cet état de fait la confirmation que ce curriculum est un référentiel d'enseignement et d'apprentissage, mais pas d'évaluation (Marc & Wirthner, 2013). Pourtant, légalement, c'est bien en référence à ses contenus que les épreuves doivent être réalisées, car elles correspondent à « la phase d'opérationnalisation du curriculum » (Marcel, 2011, p. 9). Un travail d'appropriation et d'adaptation aux situations évaluatives, fondamental, doit donc être réalisé par les praticiens. Et là, nos résultats montrent que les enseignants, pour exploiter un tel curriculum pour l'évaluation, sont amenés à mobiliser des connaissances sur les objets de savoir, à tisser des liens entre les processus d'enseignement, d'apprentissage, et les démarches d'évaluation, bref, à réaliser un travail complexe. Ce constat va dans le sens des propos de Mottier Lopez (2016), qui, à la lumière de ses travaux significatifs dans le domaine, affirme : « [pour l'enseignant] évaluer à partir du PER suppose donc, entre autres choses, de formuler des critères d'évaluation en rapport avec les objectifs visés et donc les objets de savoir en jeu, ce qui n'est pas toujours tâche facile » (p. 2). Cette orientation ne laisse toutefois pas supposer que la seule existence d'un référentiel suffirait à son exploitation par les enseignants (Mottier Lopez & Dechamboux, 2017). Mais force est de constater qu'une évaluation cohérente est tributaire d'une appropriation du curriculum par ces derniers, ce que mettent en évidence d'autres chercheurs dans des contextes différents : « Teachers need well-developed knowledge of the curriculum domain. This is foundational for effective assessment and curriculum planning and task design » (Klenowski & Wyatt Smith, 2014, p. 56). Ces éléments mettent selon nous en évidence la nécessité de penser l'articulation entre le curriculum pour l'enseignement et l'apprentissage, et le référentiel d'évaluation.

3.1.2 Les outils de notation standardisés au sein du modèle théorique

Nous avons mis en évidence que les résistances - ou les impossibilités - de plusieurs enseignants à construire des notes référées à l'apprentissage relèvent de deux problèmes majeurs : un problème important à qualifier la performance des élèves en de trop nombreux échelons (onze¹⁸⁷), et une difficulté à abandonner une logique d'échelle linéaire. Nous faisons donc ici l'hypothèse qu'une évaluation sommative notée cohérente d'un point de vue de notre modèle pourrait dès lors plus facilement exister avec des échelles plus réduites ou dans lesquelles l'insuffisance ne se traduirait que par une note¹⁸⁸.

La possibilité que le modèle de l'alignement curriculaire élargi soit un modèle théorique au service de la compréhension des pratiques évaluatives sommatives jusqu'à la notation reposerait donc en partie sur les épaules des... décideurs. En effet, les raisons pour lesquelles les enseignants sont souvent empêchés de noter en se référant aux apprentissages sont parfois plus amples et complexes que leurs incompétences (Brookhart, 2013), et renvoient aussi à des questions légales (Mc Elligott & Brookhart, 2009). Du point de vue des politiques ou des départements éducatifs, notre souhait que les directives en la matière changent semble irréaliste, tant les finalités de l'évaluation sommative à l'école sont multiples et parfois en porte-à-faux avec les pratiques réelles et les cultures d'établissement¹⁸⁹. En revanche, c'est peut-être au niveau des directions d'établissement qu'une marge de manœuvre serait souhaitable. Pour cela, les directeurs devraient eux aussi être formés à l'évaluation, posséder une réelle autonomie pédagogique, afin d'encourager les enseignants qui le demandent à se former (Guskey, 2011). Cela permettrait hypothétiquement aux praticiens de développer des pratiques d'évaluation centrées sur les apprentissages, plutôt que sur des normes extérieures aux élèves qui participent à gérer leurs flux (Brookhart, 2013).

¹⁸⁷ Pour rappel, ces onze échelons correspondent aux notes en vigueur dans le canton de Vaud, allant de 1 à 6, avec demi-notes.

¹⁸⁸ Ce qui est le cas dans les travaux que nous avons consultés et que nous citons dans ces lignes.

¹⁸⁹ Par exemple, dans le canton de Vaud, la Loi sur l'Enseignement Obligatoire (LEO) stipule à l'article 106 que l'évaluation vise à :

a. conduire l'enseignement dans le but de permettre à chaque élève d'atteindre les objectifs du plan d'études ; b. guider l'élève dans ses apprentissages ; c. dresser des bilans des connaissances et des compétences acquises, en vue des décisions de promotion, d'orientation dans les niveaux et les voies ou de certification ; d. informer les élèves et leurs parents de la progression des apprentissages.

3.1.3 Un modèle théorique « engagé » ?

Rappelons-le : le modèle théorique de l'alignement curriculaire élargi tel que nous l'avons conceptualisé poursuit une finalité compréhensive des pratiques évaluatives, et non prescriptive. Dans ce sens, nous l'avons considéré comme un outil au service du développement de pratiques cohérentes, afin que les élèves soient évalués, soutenus, accompagnés et finalement orientés sur la base de ce dont ils ont pu faire preuve, et non au regard d'autres éléments éloignés d'une logique de validation des apprentissages (Guskey et al., 2011 ; Moss, 2013 ; Winger, 2009). Certains chercheurs, comme Gauthier, Mellouki, Bissonnette et Richard (2005), ont affirmé que l'alignement curriculaire était un outil au service de la lutte contre l'échec scolaire et de l'égalité de traitement des élèves. Nous sommes aujourd'hui tenté de dire qu'il en va de même pour son acception élargie. Pour autant que les enseignants la connaissent.

Ces constats soulèvent la question des modalités de formation que nous avons mises en œuvre au sein de notre dispositif : comment avons-nous géré ces dilemmes portant sur le référentiel d'évaluation ? Quelle posture avons-nous adoptée face aux multiples interrogations et demandes des enseignants relatives aux barèmes et à certaines prescriptions ?

3.2 De quelques réflexions sur la formation au sein du dispositif

La plupart des constats relevés durant une vingtaine d'années dont nous faisons état empiriquement et de manière informelle dans notre introduction se sont confirmés dans les résultats que nous avons présentés. Qu'en déduire ? Probablement que les pratiques évaluatives sommatives sont délicates à questionner, et *a fortiori*, à amener vers davantage de cohérence (Mottier Lopez & Pasquini, 2017 ; Pasquini, 2016 ; Pasquini, 2017e). Pourtant, nous observons ici que des mouvements intéressants ont lieu chez la majorité des enseignants, parfois sur des aspects clés des pratiques, et qu'ils apparaissent assez rapidement dans le processus de formation. Pour quelles raisons ? Tentons une explication critique.

3.2.1 Un dispositif outillé conceptuellement

Le fait d'amener les enseignants à s'approprier un modèle théorique pour penser et infléchir leurs pratiques est un choix que nous avons fait dès le départ. Il nous est en effet apparu fondamental de mener ce travail à l'aide d'apports conceptuels, comme le défendent par exemple Jorro (2013) ou Altet (2000). Cette perspective a été rapidement soutenue par certains

enseignants qui, lors de notre séance 0, ont demandé à pouvoir bénéficier de tels apports, ce qui nous a permis d'accompagner l'analyse des pratiques, et surtout de modéliser leur questionnement professionnel.

Cette manière de procéder est encouragée par de nombreux chercheurs étudiant le travail en équipe (e. g., Boucenna & Charlier, 2012). Nous concevons toutefois que ce modèle a pris une place non négligeable dans les séminaires de formation ou dans les entretiens. Nous nous demandons donc dans quelle mesure la référence permanente à ce dernier n'a pas poussé parfois les enseignants à le comprendre comme étant un incontournable.

3.2.2 Un dispositif bienveillant

Cela étant, cette démarche de formation nous est apparue bienveillante pour au moins quatre raisons :

- Nous avons toujours exploité le modèle pour comprendre les phénomènes évaluatifs et non pour les normer. Les retours aux enseignants ont donc tous épousé cette logique et par là, ont participé à respecter le contrat que nous avons implicitement fixé avec les enseignants à l'issue de la séance 1 ;
- C'est aussi à l'aune de ce modèle que nous leur avons adressé les commentaires dans leurs épreuves, de manière à provoquer des débats référés aux mêmes composantes et d'un point de vue itératif ;
- Nous avons accordé une grande importance aux pratiques, notamment dans leur pouvoir de questionner le modèle. Nous avons donc toujours veillé à éviter un regard surplombant du chercheur sur les apports des enseignants ;
- Même engagées et parfois provocatrices¹⁹⁰, nos questions aux enseignants, dans leurs épreuves ou lors des séminaires de formation, ont toujours cherché à rester fidèles aux pratiques réelles et aux informations qu'ils nous ont fournies. Dans ce sens, nous avons certes orienté les enseignants vers des réflexions exigeantes, complexes, et parfois vécues comme difficiles, mais nous l'avons fait dans le respect de leurs contextes et de leur capacité d'apprentissage.

Ces quelques éléments nous confortent dans l'idée que nous avons endossé un rôle

¹⁹⁰ Nous pensons notamment au questionnement systématique relatif aux caractéristiques des objets d'apprentissage présents dans les tâches évaluatives de l'épreuve, mais surtout à celui portant sur le statut du PER ou sur l'exploitation de certains barèmes.

d'accompagnateur au sein du dispositif. Nous le développons ci-après.

3.3 De quelques réflexions sur la recherche au sein du dispositif

Nous prenons ici le parti de discuter justement notre rôle au sein de cette recherche, car c'est un des éléments centraux que nous n'avons pas encore abordé.

3.3.1 « Faire » de la recherche et « faire » de la formation

Nous l'exposons dans notre introduction de thèse et plus spécifiquement dans nos chapitres méthodologiques : c'est ici initialement un souci de formateur qui a donné lieu à une préoccupation de chercheur. Les constats du premier ont permis au second une problématisation féconde. Puis, progressivement, ce sont les données du chercheur qui ont infléchi certaines pratiques du formateur. Tout au long de la démarche, les deux ont dialogué sans cesse, avec la volonté de poursuivre conjointement un développement de connaissances pour la pratique et pour la recherche (Desgagné et al., 2001). En effet, il s'est agi notamment pour nous de construire un dispositif pour former des enseignants en questionnement, mais aussi pour récolter des données permettant possiblement de répondre à des questions de recherche. Dans ce sens, nous avons été constamment habité par un souci d'accompagner les enseignants, tout en restant proche de nos interrogations scientifiques.

Nous faisons l'hypothèse que ce travail articulé a été possible car aussi bien le formateur que le chercheur que nous avons été ont pensé les pratiques évaluatives au travers du même modèle. Plus spécifiquement, le formateur l'a exploité pour comprendre, interroger, commenter les pratiques des enseignants et les accompagner dans leurs apprentissages, alors que le chercheur l'a utilisé comme outil de catégorisation de sens pour interpréter ses diverses données. Toutefois, on peut se demander dans quelle mesure ces deux rôles, bien qu'articulés dans les séminaires de formation de manière assumée, ont réellement été perçus par les enseignants comme étant bien distincts. Une des manières de répondre à cette interrogation est de donner à voir dans quelle mesure le rôle d'accompagnateur que nous avons pensé investir a pu réellement prendre forme.

3.3.2 L'accompagnateur

L'acte d'accompagner sous-tend l'idée du chemin, du processus. Il se destine par ailleurs à autrui : on accompagne toujours quelqu'un pour l'amener quelque part. En formation continue,

l'accompagnement se réalise auprès de personnes qui possèdent un but, qui sont ouvertes à ajuster leurs pratiques (Savoie-Zajc, 2010). Dans cette perspective, il poursuit une finalité de développement professionnel. Par ailleurs, l'accompagnement peut prendre des formes diverses mais singulières, et c'est ce qui fait sa richesse (Charlier & Biemar, 2012). A ce titre, nous avons non seulement guidé le travail réflexif des enseignants, mais avons aussi offert des apports théoriques pour que ces derniers comprennent mieux les phénomènes qui leur posaient problème ou qui les interrogeaient. Cette manière d'être avec autrui s'est incarnée dans un double processus : « l'accompagnateur pose les conditions pour que l'Autre fasse son chemin et ainsi construit avec lui son chemin » (Vial, 2007, p. 34-35).

L'accompagnement ainsi thématisé repose donc sur une relation interpersonnelle de confiance, considère le sujet comme étant dépositaire de connaissances importantes relatives à son développement, et requiert une pratique réflexive tant pour l'accompagné que pour l'accompagnateur. Dès lors, une des compétences clés de ce dernier est l'analyse de pratiques (Donnay & Charlier, 2008), ce qui rejoint un des aspects de notre dispositif.

Certains chercheurs ayant documenté l'accompagnement d'équipes d'enseignants (e. g., Cuche, Canivet & Donnay, 2012) montrent que l'accompagnateur adopte plusieurs postures qui définissent des champs de compétences à développer. Nous les reprenons ici en les mettant en lien avec notre propre travail, car ils mettent en évidence les champs que nous avons investis. Ainsi, comme le précisent Charlier et Biémar (2012, pp.158-159) ;

- *L'accompagnateur est un praticien réflexif.* Cette réflexivité lui permet de se distancer d'une posture surplombante et l'engage dans une analyse permanente de sa propre pratique, ici de formateur, mais aussi de chercheur. Nous retrouvons là une des caractéristiques du chercheur au sein d'une méthode collaborative (Desgagné, 1997 ; Van Nieuwenhoven & Colognesi, 2015).
- *L'accompagnateur est gardien du cap.* Il thématise et se réfère en permanence aux objectifs poursuivis, ce que nous avons régulièrement fait.
- *L'accompagnateur est garant des cohérences du dispositif.* Il cadre les échanges et encourage le respect des opinions et choix de chacun. Mais il est également responsable de la cohérence théorique et méthodologique des approches qu'il propose, comme ici.
- *L'accompagnateur est régulateur de développement.* La démarche étant processuelle et collaborative, il respecte et assure que chacun va pouvoir progresser en fonction de ses

questions et connaissances. Cette posture renvoie aux conditions nécessaires à un développement professionnel que nous débattons au point 7 de ce chapitre.

- *L'accompagnateur est un catalyseur.* Il modélise les situations travaillées, favorise l'appropriation des contenus par les participants. Là, nous avons assumé ce rôle.
- *L'accompagnateur est un référent à un tiers objet.* Notre posture d'extériorité nous a autorisé, voire encouragé à exploiter des modèles théoriques pour analyser les situations qui questionnent.

Comme nous en avons fait l'hypothèse dans notre chapitre méthodologique (partie 6), l'accompagnateur occupe ici bel et bien un triple rôle : chercheur, formateur d'enseignants et accompagnateur d'équipes d'enseignants mobilisés par une même problématique (Boucenna & Charlier, 2012). Notre travail d'accompagnement, centré notamment sur l'appropriation de connaissances et le développement de compétences (Paquay, Van Nieuwenhoven & Wouters, 2010 ; Popham, 2009, 2011), mais aussi porté par des préoccupations de recherche, a donc poursuivi une double finalité de développement professionnel : pour les personnes avec lesquelles il travaille, et pour lui-même.

L'accompagnateur que nous avons été est dès lors un acteur-clé de toute démarche de formation saisie dans un processus de recherche à visée de collaboration. Nous avons, il nous semble, participé à intensifier un travail collaboratif entre des enseignants, un formateur et un chercheur, et cela même si nous avons été ce formateur et ce chercheur. Ce constat va dans le sens de bon nombre de travaux s'accordant sur le fait que travailler ensemble, et de surcroît dans une perspective collaborative et d'accompagnement, concourt à la construction et à la production de développement professionnel pour tous les acteurs concernés (e. g., Letor & Périsset Bagnoud, 2010)¹⁹¹.

Cela étant, en quoi sa conceptualisation approfondie, réalisée en amont du processus de recherche-formation, nous aurait-elle permis de réduire quelques biais liés à la complexité de notre posture ? Ce questionnement ouvre à présent une discussion des limites de notre recherche.

¹⁹¹ Pour les travaux mentionnés, se référer à l'ouvrage coordonné par Lise Corriveau et al. (2010) : Travailler ensemble dans les établissements scolaires et de formation, chez De Boeck Supérieur.

4 Les limites de notre recherche

Nous développerons cinq limites en rapport étroit avec notre problématique. Les trois premières concernent les aspects conceptuels de notre travail et sont en lien. Les deux suivantes sont davantage en rapport avec des questions méthodologiques.

4.1 Nous pensions traiter de validité en évaluation, mais...

Au départ, nous avons pensé que la validité serait au cœur de nos préoccupations de chercheur, puisqu'elle est centrale dans la compréhension des pratiques évaluatives. Nous avons documenté le fait que, pour certains chercheurs, la validité est un des indicateurs principaux permettant de juger de la qualité des pratiques évaluatives sommatives (Biggs, 2003 ; Brookhart, 2003 ; Klenowski & Wyatt-Smith, 2014). Il nous a semblé que des questions comme « le contenu de la tâche évalue-t-il ce que l'évaluateur pense évaluer ? », « les critères sont-ils formulés en référence au contenu et au niveau de complexité des objectifs évalués ? » ou « la note rend-elle compte des apprentissages réussis et de ceux qu'il est encore nécessaire d'approfondir au regard des processus d'enseignement et d'apprentissage ? » étaient importantes dans la problématisation de notre situation de recherche, car elles traduisent cette quête de validité.

Bonner (2013) conforte ce questionnement, lorsqu'elle explicite les cinq caractéristiques que devrait posséder toute démarche d'évaluation valide. En effet, la première consiste bel et bien en une forte cohérence d'alignement curriculaire entre les processus d'enseignement, d'apprentissage et d'évaluation, ce que nous avons investigué en grande partie. Mais elle va plus loin, en précisant encore que la validité dépend également de la résolution maximale des biais durant le processus évaluatif, de la possibilité d'évaluer non seulement les produits mais surtout les processus, du fait de juger de la pertinence des résultats obtenus, et enfin de la transparence des démarches perceptible pour tous les acteurs concernés.

En dépouillant nos données, puis en les analysant, il nous est donc apparu que cette problématique de la validité était partiellement saisie dans la recherche de cohérence curriculaire induite de manière continue et permanente par notre modèle théorique (Hammerness, 2006 ; Martone & Sireci, 2009). Aussi avons-nous abandonné assez rapidement la catégorie conceptualisante « validité » dans nos premiers résultats, car elle ne donnait rien à comprendre de plus que ce que d'autres catégories ne montraient ou n'explicitaient déjà.

Ce choix représente vraisemblablement une limite de notre recherche, même si notre modèle nous a permis de traiter indirectement cette question au travers notamment de la cohérence des pratiques (Biggs, 2003 ; Bateman et al., 2009). Cela ne signifie en rien toutefois que cohérence et validité sont de stricts synonymes, bien au contraire. Les deux notions, comme nous le montrons, sont conceptualisées de manière contrastée, notamment par le fait que la validité ne dépend pas d'un modèle théorique défini. C'est peut-être à ce niveau que nous aurions pu être plus précis.

4.2 Nous pensions thématiser un jugement professionnel holistique, mais...

Le jugement évaluatif est lui aussi omniprésent dans notre recherche collaborative (Allal & Mottier Lopez, 2014). Nous avons tenté de montrer sur la base de quelles informations il se fonde au cas par cas chez les enseignants, en mettant en exergue ses diverses composantes. Nous avons mis en évidence le fait que, progressivement, les praticiens changent les références¹⁹² à partir desquelles ils formulent leurs décisions évaluatives. Leurs pratiques sont remises en question, leurs croyances évoluent : autant de signes, pour Laveault (2008), que « le jugement devient professionnel » (p. 487), ce qui traduirait une reconstruction critique de la pratique (Coles, 2002).

Toutefois, la conceptualisation de ce jugement est lacunaire. En effet, si nous avons émis l'hypothèse qu'un jugement professionnel holistique peut possiblement voir le jour chez les enseignants lorsqu'ils sélectionnent les informations pour prendre leurs décisions évaluatives à l'aune des diverses composantes du modèle théorique, comme c'est le cas ici, nous ne sommes pas allé jusqu'à modéliser ce jugement. Il serait donc pertinent selon nous, dans des recherches à venir, de donner à voir à quelles conditions ce jugement, ancré dans une recherche de cohérence au niveau des objets de savoir à faire apprendre et à évaluer (dimension curriculaire), et en permanence référé à des critères contextualisés à la situation d'évaluation et possiblement progressifs (dimension holistique), pourrait exister, et cela même lorsque les pratiques de notation ou de pondération ne sont pas à proprement parler holistiques.

4.3 Au-delà de l'alignement curriculaire élargi, point de salut ?

Souhaitons que le lecteur ne quitte pas cette thèse en ayant le sentiment d'avoir assisté à un

¹⁹² Dans leurs travaux, Lucie Mottier Lopez et Gérard Figari parlent de référents.

éloge sans contrepoint du modèle. Il nous semble l'avoir mis en débat à plusieurs reprises, notamment lorsque nous l'avons soumis aux contraintes de la pratique évaluative réelle. Ses atouts, nombreux, ont certes été mis en évidence, mais également ses limites. Toutefois, et même si nous continuons à penser que la recherche et la formation possèdent ici un outil théorique puissant pour comprendre et modéliser les pratiques évaluatives sommatives, on pourrait nous reprocher de ne pas avoir tenté de le soumettre à certaines perspectives didactiques, ou à des conceptions de l'apprentissage et de l'enseignement largement partagées dans le monde de l'éducation et de la formation.

Cette réflexion nous semble tout à fait pertinente. En effet, il aurait été judicieux, par exemple, de confronter notre modèle théorique à l'approche par compétences et aux multiples recherches qui l'ont traitée (e. g., Beckers, 2011 ; Carette, 2007 ; De Ketele & Gérard, 2005 ; Le Boterf, 2001 ; Gérard, 2009 ; Legendre, 2008 ; Rey, Carette, Defrance & Kahn, 2012 ; Scallon, 2007) - ou de donner à voir dans quelle mesure cette approche aurait interagi avec le modèle - et ceci d'autant plus avec un plan d'études formulé en ces termes (Marc & Wirthner, 2013)¹⁹³. Dans le même sens, articuler des apports didactiques sur les objets de savoir avec le modèle aurait vraisemblablement permis d'aller plus loin dans l'interprétation de certaines données, et possiblement dans l'investigation des résistances rencontrées dans certaines pratiques évaluatives. Deux champs de recherche au moins demandent donc potentiellement à être davantage investis.

4.4 De quelques biais méthodologiques

Selon nous, trois biais principaux sont à l'œuvre dans notre recherche. Commençons par ceux qui concernent le chercheur.

Un biais interprétatif

Un risque de biais interprétatif est bien présent dans cette recherche, surtout dans la réalisation des portraits où l'analyse doit prendre en compte un nombre important de données à trianguler (Silverman, 2009). Nous avons travaillé abondamment à ne garder que les données les plus représentatives, relativement à nos questions de recherche, mais nous sommes conscient que le risque « de faire dire aux données ce que l'on croit qu'elles disent » existe bel et bien. Ce biais

¹⁹³ Avec cette limite, nous nuancions nos propos émis sur l'approche par compétences au point 12 de notre revue de littérature (Partie 2).

interprétatif pourrait également provenir de la proximité que nous entretenons avec l'objet de recherche. De manière complémentaire à ce que nous affirmons dans l'introduction, travailler un objet que l'on côtoie depuis longtemps en tant que formateur peut également créer de fortes attentes. Nous avons dû alors redoubler de vigilance et nous doter d'outils méthodologiques permettant d'aller vers la plus grande validité possible. Toutefois, le risque de sur-interpréter des données est bien présent.

Un biais de désirabilité sociale

Toute recherche est aux prises avec ce biais, naturel, omniprésent, et l'important est de le considérer comme pouvant potentiellement invalider certaines démarches ou résultats. Cela étant, dans quelle mesure a-t-il pu influencer les enseignants ayant participé à la recherche ? Peut-on considérer qu'une partie des mouvements dans les pratiques, des propos, des écrits documentés dans notre travail seraient le fruit d'une conformité recherchée par les enseignants afin d'être compatibles avec des attentes ? Il nous semble avoir pris des précautions méthodologiques pour diminuer ce biais, notamment lorsque nous avons demandé aux enseignants de produire des épreuves pour leurs classes, ou que nous les avons confrontés aux pratiques de leurs pairs. Le but était par là de les ramener dans leurs contextes, dans leurs préoccupations propres, et de les décentrer des attentes pouvant être perçues chez l'accompagnateur. Toutefois, le doute persiste.

Un biais sur l'objet

Nous avons récolté des traces des pratiques des enseignants au travers des épreuves. Cela demeure toutefois des traces indirectes à partir desquelles les participants ont produit un discours (oral, écrit). Nous n'avons pas récolté de données lorsqu'ils élaboraient leurs épreuves, ni quand ils les corrigeaient. Nous n'avons pas non plus observé la manière avec laquelle ils faisaient passer ces épreuves aux élèves, ni comment ils préparaient ces derniers, au quotidien, au travers de leur enseignement et des situations d'apprentissages proposées. Nous faisons dès lors l'hypothèse que des écarts ont pu exister entre ce que les enseignants ont affirmé ou donné à voir, et ce qu'ils ont réellement fait. Il aurait ici été pertinent d'interroger ces décalages (Perrenoud, 2002). Par ailleurs, cette lacune peut être comprise à juste titre comme contradictoire avec certaines caractéristiques de notre modèle théorique qui met les processus d'enseignement et d'apprentissage au centre de sa conceptualisation. Dans le même sens, nous n'avons pas non plus investigué les conceptions de l'apprentissage des participants, alors

qu'elles semblent jouer un rôle fondamental dans le développement des compétences évaluatives (Timperley, 2011).

4.5 Que sont devenus les enseignants de la recherche ?

On pourrait également discuter la durée de notre recherche et se demander pourquoi nous ne sommes pas retourné vers les participants pour voir comment ils ont exploité les apports de la formation : que leur reste-t-il, en termes de pratiques, de réflexions, de points d'attention ? Quelles dimensions des pratiques ont résisté aux contextes, aux habitudes, aux prescriptions, aux habitus ? Lesquelles ont cédé, et pourquoi ? Lesquelles ont participé à transformer certains de leurs aspects environnementaux ? Autant d'interrogations qui restent en suspens, et qui, vraisemblablement, nous titilleront encore longtemps. En effet, bien que longitudinale¹⁹⁴, collaborative et centrée sur les pratiques réelles, cette recherche reste un arrêt sur image. Même si nous collaborons encore avec trois participants à différents niveaux, nos démarches avec eux restent informelles.

Ces limites, non exhaustives, nous poussent maintenant à penser à de nouvelles perspectives larges pour la recherche et pour la formation.

5 Quelques perspectives pour des recherches futures

Dans le premier chapitre du *Handbook of Research on Classroom Assessment* (CA) publié en 2013 et faisant pour nous référence, McMillan affirme : « There is much to be done to establish a body of knowledge about CA » (p. 12). Dans ce sens, trois perspectives de recherche méritent selon nous d'être investiguées suite à ce travail. Elles devraient idéalement, comme nous l'avons évoqué plus haut, associer des chercheurs en évaluation et des spécialistes des didactiques, l'influence des objets de savoir en évaluation étant significative. Nous plaidons à ce niveau et fidèlement à notre modèle théorique, comme Robbes (2013), pour « un dépassement de l'opposition entre didactique et pédagogie » (p. 7).

5.1 Considérer la notation comme un objet de recherche à part entière

Ce travail de thèse nous a permis de constater que la recherche en évaluation francophone n'a

¹⁹⁴ Rappelons que neuf mois se sont déroulés entre l'entretien de recrutement et le dernier entretien de recherche. « Le temps d'une grossesse », comme l'a dit humoristiquement Samia, faisant allusion à son temps de gestation pour appréhender les contenus de la formation.

pas développé de concept, de modèle, ni de champ propre à la notation en général - le « *grading* » -, et encore moins lorsqu'elle est envisagée comme pouvant soutenir les apprentissages. Pourtant, certains chercheurs, comme Hadji (2017), défendent aujourd'hui l'idée que « l'attribution d'une note présuppose toute une série de jugements préalables. Toute la question pour l'évaluateur sera de savoir si une note a la capacité d'exprimer adéquatement la spécificité du jugement d'acceptabilité qu'il a été amené à formuler » (p. 245). La question de la cohérence de la note est ici bien présente. Dès lors, riches de notre connaissance approfondie des enjeux qui gravitent autour d'elle (Butera et al., 2010 ; Merle, 2007 ; Randall & Engelhard, 2010), de notre conscience que les pratiques de notation sont fréquemment idiosyncratiques, fort peu outillées, et sachant que malgré une attention des enseignants à noter le plus possible en référence aux apprentissages (Walvoord & Johnson Anderson, 2009), les notes manquent souvent de cohérence (Black et al., 2010 ; Winger, 2009), nous devons investiguer de nouvelles perspectives de recherches sur le sujet.

En effet, la note ne doit plus être assimilée ni réduite au résultat d'une transaction techniciste (Hadji, 2016), mais bien être comprise comme un « signe » dont l'importance est fondamentale dans les processus de communication entre les enseignants et les élèves (Mottier Lopez & Allal, 2008). Noter est définitivement une pratique complexe de la profession (Lenoir, 2004). Et, comme Brookhart (2013), nous considérons que ce n'est pas le recours aux évaluations externes qui, à terme, remplira la fonction de certification des apprentissages.

Il est donc important que la communauté scientifique francophone s'empare plus intensément de cet objet qu'est la notation, dont les premières études, à notre connaissance, datent de 1888 avec les travaux d'Edgeworth aux Etats-Unis (Brookhart, Guskey, Bowers, McMillan, Smith, & Smith, (...), 2016) et ceux de Piéron en France en 1922¹⁹⁵. Des chercheuses comme Mottier Lopez (2009, 2015), Allal (2010) ou encore Harlen (2005 ; 2012) ont ouvert un sillon, en encourageant à questionner les frontières conçues comme hermétiques entre évaluation formative et évaluation sommative. Notre souhait d'approfondir une notation à visée de soutien à l'apprentissage s'inscrit pleinement dans cette préoccupation scientifique, notamment parce que les élèves méritent des retours sur leurs acquis et leurs lacunes au-delà des chiffres : « the commentary that accompanies a grade can mean a lot to a student » (Smith, Smith & De Lisi, 2001, p. 105). En d'autres termes, il s'agit de se demander à quelles conditions les informations

¹⁹⁵ Les résultats de cette première étude ne seront publiés qu'en 1927 dans la revue « L'Année psychologique » (t. XXIII, pp. 146-175).

délivrées par la note peuvent poursuivre une visée formative (Good, 2011 ; Laveault & Allal, 2016). Par exemple, Brookhart (2017), dont le titre du dernier ouvrage est tout un symbole¹⁹⁶, montre qu'à l'aide d'outils conceptuels, il n'est plus illusoire de penser que les pratiques évaluatives notées peuvent évoluer vers davantage de cohérence et fournir des feedbacks centrés sur les apprentissages, en vue de les soutenir. Des perspectives prometteuses semblent selon elle se dessiner : « as I look over the history of grading [...] it seems to me that we may actually be witnessing a genuinely new development » (p. 143). Il y a donc une nécessité d'approfondir et de développer des modèles théoriques pour investiguer cette problématique (Brookhart, 2013). La recherche francophone, davantage portée sur l'étude sociologique des phénomènes relatifs à la notation, serait par là à même de mutualiser ses savoirs avec ceux de la communauté anglophone, plus pragmatiques et relevant des pratiques évaluatives. Il y a fort à parier que les enseignants de Grande Bretagne, des Etats-Unis, de Nouvelle-Zélande ou de Scandinavie se posent les mêmes questions que leurs homologues français, belges ou suisses romands lorsqu'il s'agit de certifier et de noter¹⁹⁷. Souhaitons qu'il en aille bientôt de même pour les chercheurs. Notre modeste et présent travail va dans ce sens. Il s'agit maintenant d'aller plus loin, en mettant en oeuvre, par exemple, des projets de recherche comparatistes ou collaboratifs partant de l'étude des pratiques réelles.

5.2 Partir des pratiques réelles et contextualisées

Nous avons montré dans notre revue de littérature que les recherches en évaluation sur les pratiques déclarées sont foison (e. g. Braxmeyer et al., 2005 ; Moss, 2013 ; Brookhart, 2013). Les travaux sur les pratiques réelles sont quant à eux bien plus confidentiels, surtout en évaluation sommative, traduisant un manque certain à ce niveau. Ces constats devraient selon nous encourager les chercheurs à privilégier cette dernière orientation de recherche, car le contexte des pratiques a une influence majeure sur leur possible développement, et donc, en conséquence, sur le développement professionnel des praticiens¹⁹⁸ (Popham, 2009, 2011 ; DeLuca & Johnson, 2017). Là, les chercheurs pourraient partir du questionnement des enseignants, faisant l'hypothèse qu'il est garant de leur engagement et du sens qu'ils peuvent trouver à s'investir dans une démarche de recherche-formation (Laveault, 2010 ; Paquay, 2000 ;

¹⁹⁶ How to use grading to improve learning ?

¹⁹⁷ Cette préoccupation commune sur l'évaluation est par exemple très présente dans le numéro spécial de mai 2017 de la revue « *Assessment in Education* » intitulé *Developing Teachers' Assessment Capacity*, qui réunit des articles de chercheurs du monde entier et auquel Lucie Mottier Lopez et nous-même avons participé.

¹⁹⁸ Nous justifions ce lien au dernier point de cette partie.

Van Nieuwehoven & Colognesi, 2015)¹⁹⁹.

Certes, observer des pratiques au travers d'un dispositif de formation ou les recenser *via* un questionnaire sont deux démarches de recherche complémentaires qui poursuivent des buts différents. Toutefois, travailler *à partir* des pratiques des enseignants et non *sur* leurs pratiques comporte potentiellement plus d'obstacles (logistiques, matériels, contextuels, financiers, administratifs). Le risque de voir des enseignants se désengager du projet de recherche est également plus grand, puisque leur investissement est coûteux en énergie et en remises en question²⁰⁰. Mais nous pensons qu'il faut persévérer dans ce sens. D'une part, car les connaissances de la recherche à ce niveau sont encore insuffisantes, et d'autre part parce que celles dont nous disposons tendent à montrer que les recherches sur les pratiques effectives donnent de meilleurs indicateurs sur le développement des compétences des enseignants que les enquêtes (Campbell, 2013).

5.3 Penser le lien entre recherches collaboratives et communautés de pratiques

Notre recherche collaborative a duré six mois. Puis chacun est retourné dans sa classe, dans son environnement. Cette rupture entre le temps où des enseignants et un chercheur collaborent étroitement et le moment où ils se séparent nous interpelle. Ne devient-il pas impératif que les enseignants, les chercheurs et les formateurs travaillent ensemble sur le long terme, dans des groupes de recherche-formation au sein desquels les premiers pourraient notamment partager leurs questionnements et mettre à l'épreuve de connaissances scientifiques leurs pratiques en vue de les optimiser ou mieux les comprendre, et les seconds, par exemple, aller vers le terrain en mettant en discussion leurs approches théoriques ? Pour inscrire de tels projets dans des durées plus conséquentes, ne devrait-on pas penser à développer davantage des communautés d'apprentissage, comme le suggèrent notamment Fullan (2001) ou Savoie-Zajc (2010)²⁰¹ ? Cette question dépasse vraisemblablement les acteurs concernés en priorité, puisqu'elle engage

¹⁹⁹ Ce sont par ailleurs ces mêmes questions qui ont été exposées et négociées dans la phase de cosituation de notre dispositif de recherche collaborative.

²⁰⁰ Rappelons que Samia, fort déstabilisée, a failli quitter le projet de recherche suite au deuxième séminaire de formation. Nous ne l'avons appris qu'à la fin du dispositif.

²⁰¹ Cette dernière auteure, reprenant notamment les travaux d'Orellana (2002), définit les éléments essentiels et constitutifs d'une communauté d'apprentissage : « la présence d'un projet, d'une intention de développement, d'une recherche de croissance commune au sein d'un groupe » (p. 14). Nous retrouvons ici les finalités de l'accompagnement explicitées plus haut. Par ailleurs, Laveault (2008) montre que ces communautés permettent d'« unir les compétences de plusieurs enseignants dans la prise de décision » (p. 495) et ainsi optimiser leur jugement professionnel.

en amont la responsabilité et les choix des décideurs, des chefs d'établissement, ou encore des directeurs d'institution de formation, sans l'aval desquels le terrain ne peut être un lieu d'expériences professionnelles (Snoeckx, 2008). A ce niveau, nous faisons l'hypothèse que la collaboration « interroge fortement les conditions institutionnelles à mettre en place pour que les enseignants voient l'intérêt de s'inscrire dans des démarches collaboratives avec des chercheurs » (Vinatier & Morrisette, 2015, p. 162).

Nous sommes riches de perspectives pour la recherche. Mais le lecteur ne sera pas surpris de constater notre impossibilité à aller de l'avant sans énoncer quelques perspectives possibles également d'un point de vue de formation.

6 Quelques perspectives pour la formation en évaluation sommative notée

Deux perspectives distinctes nous semblent ici à mettre en débat, en fonction des différents publics potentiels auxquels elles peuvent respectivement s'adresser.

6.1 Construire la formation continue en référence aux « questions vives de la pratique »

Nous postulons que de nombreux enseignants, tous cycles confondus, sont ouverts à se former au niveau de leurs pratiques évaluatives sommatives lorsque leur questionnement est pris au sérieux et fait partie intégrante de la formation qui leur est proposée. Un des enjeux majeurs de telles actions de formation réside dans le fait d'accéder à ces « questions vives de la pratique » (Pasquini 2017a, 2017b, 2017c, 2017d, 2017e).

Il serait dès lors important que tout institut de formation encourage la création d'équipes de formateurs non seulement experts au niveau des pratiques évaluatives sommatives notées, mais aussi de l'accompagnement des équipes, de manière à intervenir dans les établissements sur ces questions (Paquay, 2012). Dans le même sens, les chefs d'établissement, encore une fois, devraient idéalement être sensibilisés aux dimensions complexes de cette pratique pour assumer un leadership pédagogique (Pasquini, 2000 ; Pelage, 2003) et éviter, entre autres, de formuler des sur-prescriptions bridantes et a-pédagogiques, ou pire encore, comme nous l'avons vécu, de bloquer l'accès de leurs enseignants à des formations portant sur le sujet. L'innovation des pratiques est à ce prix (Blanchet, 2010). La formation dans ce champ concerne donc tous les acteurs du système, dès la formation initiale.

6.2 Concevoir la formation initiale autour de connaissances clés

A la lumière des nombreux travaux que nous avons étudiés, il n'existe pas de consensus sur les connaissances que les enseignants devraient posséder pour évaluer et noter leurs élèves de manière cohérente. Pourtant, de multiples sources sur l'*assessment literacy* mettent tout de même en évidence le fait que certaines d'entre elles semblent incontournables (Xu & Brown, 2016).

A nos yeux, ce constat devrait encourager les instituts de formation à approfondir la réflexion sur cette question des connaissances en évaluation, à l'aune des enjeux et des contraintes que les futurs enseignants vont devoir affronter²⁰². A ce titre, s'il semble partagé que des connaissances sur les dimensions larges de l'évaluation sont essentielles (les finalités, les fonctions, la philosophie des différentes démarches), il apparaît qu'elles sont aussi très insuffisantes pour outiller pragmatiquement les futurs enseignants (DeLuca & Klinger, 2010 ; Mertler, 2004, 2009). Des connaissances spécifiques et contextualisées permettant de modéliser les pratiques sont donc requises : celles qui sont en jeu dans notre recherche pourraient constituer un point de départ du travail à mener.

D'un point de vue de formation initiale, développer des connaissances et des compétences en évaluation sommative devrait donc selon nous être une priorité. Nous pensons même que la formation à l'évaluation peut, à certaines conditions, permettre une meilleure compréhension des enjeux relatifs à l'apprentissage et à l'enseignement. En d'autres termes, et en suivant plusieurs chercheurs comme Klenowski et Wyatt-Smith (2014) ou Guskey et al. (2011), nous faisons l'hypothèse qu'il est possible de penser l'enseignement et l'apprentissage depuis l'évaluation, et non seulement l'inverse.

7 Une recherche, *in fine*, pour le développement professionnel de tous les acteurs ?

La méthode de recherche collaborative vise le développement professionnel de tous les acteurs, dans la mesure où elle valorise la production de connaissances pour la recherche et pour la pratique (Van Nieuwenhoven & Colognesi, 2015). Par ailleurs, il existe un enjeu de développement professionnel très marqué dans la perspective théorique de l'*assessment literacy*, puisqu'elle favorise l'appropriation de connaissances et le développement des

²⁰² Cette réflexion vaut également pour la formation continue.

compétences chez les enseignants (Popham, 2009, 2011 ; Edwards, 2017). Un même enjeu est présent dans l'accompagnement d'équipes d'enseignants autour de questions communes et partagées. Dès lors, dans quelle mesure les participants à notre étude ont-ils entamé un processus de développement professionnel ? Nous ne dépasserons évidemment pas le stade des hypothèses pour tenter de répondre à cette interrogation, l'orientation de notre recherche n'ayant pas suffisamment conceptualisé cet objet depuis le départ.

Uwamariya et Mukamurera (2005), suite à leur revue de littérature sur le concept, définissent le développement professionnel comme « un processus de changement, de transformation, par lequel les enseignants parviennent peu à peu à améliorer leur pratique, à maîtriser leur travail et à se sentir à l'aise dans leur pratique » (p. 148). Au vu de nos résultats et de nos choix, nous pourrions affirmer que les participants ont dès lors engagé un processus de changement allant dans ce sens. Mais étant donné que la recherche collaborative devrait avoir à ce niveau comme finalité « le développement professionnel de chaque catégorie d'acteurs concernés » (Van Nieuwenhoven & Colognesi, 2015, p. 105), une réflexion serait donc également à développer au niveau de l'accompagnateur : avons-nous pu, de notre côté, augmenter notre expertise des objets de recherche et de formation que nous avons abordés ? Dans quelle mesure avons-nous modifié nos manières de procéder au cours du dispositif ? En quoi nos méthodes de chercheur et nos actions de formateur ont-elles pu bénéficier des apports du présent travail ? En d'autres termes, nous sommes-nous également développé professionnellement ?

La rigueur scientifique nous impose de suspendre tout essai de réponse. Mais une suite à notre étude pourrait se situer, déjà, là. Dans le même élan, nous constatons que notre motivation à penser des recherches futures en évaluation sommative avec des enseignants est intacte. Avec cette thèse, notre curiosité s'est accrue. De nouvelles questions ont émergé. Des incomplétudes se sont révélées. Des projets sont déjà, informellement, en construction. Tant d'objets demandent encore à être investigués. Ils sont là. Ils (nous) attendent.

Partie 10 - Bibliographie

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20 (1), 91-106.

Airasian, W. & Miranda, H. (2002). The Role of Assessment in the Revised Taxonomy. *Theory into Practice*, 41 (4), 249-254.

Alkin, M. C. (2003). Evaluation Theory and Practice : Insights and New Directions. *New directions for evaluation*, 97, 81-89.

Allal, L. (2002). Acquisition et évaluation des compétences en situation scolaire. In J. Dolz & E. Ollagnier (Eds.), *L'énigme de la compétence en éducation* (pp. 77-94). Bruxelles : De Boeck.

Allal, L. (2008). Évaluation des apprentissages. In A. van Zanten (Dir.), *Dictionnaire de l'éducation* (pp. 311-314). Paris : Presses universitaires de France.

Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire. In L. Mottier Lopez & M. Crahay (Eds.), *Evaluations en tension* (pp. 29-45). Bruxelles : De Boeck.

Allal, L. (2010). Assessment and the Regulation of Learning. *International Encyclopedia of Education*, 3, 348-352.

Allal, L. (2012). Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence. In L. Mottier Lopez & G. Figari (Eds.), *Modélisations de l'évaluation en éducation* (pp. 181-194). Bruxelles : De Boeck.

Allal, L. & Mottier Lopez, L. (Ed.) (2007). *Régulations des apprentissages en situation scolaire et en formation*. Bruxelles : De Boeck.

Allal, L. & Mottier Lopez, L. (2008). Mieux comprendre le jugement professionnel en évaluation: apports et implications de l'étude genevoise. In L. Lafortune & L. Allal (Éd.), *Jugement professionnel en évaluation: pratiques enseignantes au Québec et à Genève* (pp. 223-239). Québec: Presses de l'Université du Québec.

Allal, L. & Mottier Lopez, L. (2014). Teachers' professional judgment in the context of collaborative assessment practice. In C. Wyatt-Smith & V. Klenowski (Ed.), *The enabling power of assessment* (pp. 151-165). London : Springer.

Allen, J. D. (2005). Grades as Valid Measures of Academic Achievement of Classroom Learning. *The Clearing House*, 78 (5), 218-223.

Altet M. (2000). L'analyse de pratiques, une démarche de formation professionnalisante ? *Recherche et Formation*, 35, 25-41.

Amigues, R. & Zerbato-Poudou, M.-T. (1996). *Les pratiques scolaires d'apprentissage et d'évaluation*. Paris : Dunod.

Anadon, M. & Savoie-Zajc, L. (2007). La recherche-action dans certains pays anglo-saxons et latino-américains : une forme de recherche participative. In M. Anadon (Ed.), *La recherche participative: multiples regards* (pp. 13-30). Québec: PUQ.

Anadon, M. & Guillemette, F. (2007). La recherche qualitative est-elle nécessairement inductive ? *Recherches qualitatives*, 5, 26-37.

Anderson, L. W. (2002). Curricular Alignment : A Re-Examination. *Theory into Practice*, 41 (4), 255-260.

Anderson, L. W. & Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York : Longman.

Andrade, H. G. (2005). Teaching With Rubrics: The Good, the Bad, and the Ugly. *College Teaching*, 53, (1), 27-31.

Antibi, A. (2003). *La constante macabre ou comment a-t-on découragé des générations d'élèves*. Toulouse : Math'Adore.

Bain, D. (2002). De l'évaluation aux compétences : mise en perspective de pratiques enseignantes. In J. Dolz & E. Ollagnier (Eds.), *L'énigme de la compétence en éducation* (pp. 129-145). Bruxelles : De Boeck.

- Balach, C. A & Szymanski, G. J. (2003, avril). *The growth of a professional learning community through collaborative action research*. Conférence présentée à l'Annual Meeting of the American Educational Research Association, Chicago. Repéré à : <http://files.eric.ed.gov/fulltext/ED477527.pdf>
- Baribeau, C. & Royer, C. (2012). L'entretien individuel en recherche qualitative : usages et modes de présentation dans la Revue des sciences de l'éducation. *Revue des sciences de l'éducation*, 38 (1), 23-45.
- Barlow, M. (2003). *L'évaluation scolaire, mythes et réalités*. Paris: ESF.
- Barry, S. & Saboya, M. (2015). Un éclairage sur l'étape de co-situation de la recherche collaborative à travers une analyse comparative de deux études en didactique des mathématiques. *Recherches qualitatives*, 34 (1), 49-73.
- Bateman, D., Taylor, S., Janik, E. & Logan, A. (2009). Curriculum coherence and student success. *Pédagogie collégiale*, 22 (5), 8-18.
- Beckers, J. (2011, 2è Ed.). *Développer et évaluer des compétences à l'école: vers plus d'efficacité et d'équité*. Bruxelles : Labor Education.
- Bednarz, N. (2013). *Recherche collaborative et pratique enseignante. Regarder ensemble autrement*. Paris : L'Harmattan.
- Bednarz, N. (2015). La recherche collaborative. *Carrefours de l'éducation*, 39 (1), 171-184.
- Berger, E., Crescentini, A., Galeandro, C. & Crohas, G. M. (2010, septembre). *La triangulation au service de la recherche en éducation*. Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF). Université de Genève : Genève. Repéré à : <https://plone.unige.ch/aref2010/communications-oraales/premiers-auteurs-en-b/La%20triangulation.pdf>.
- Biberman-Shalev, L., Sabbagh, C., Resh, N. & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education*, 27, 831-840.
- Biggs, J. (1999). What the Student Does: teaching for enhanced learning. *Higher Education*

Research & Development, 18 (1), 57-75.

Biggs, J. (2003). *Aligning teaching for Constructive Learning*. The Higher Education Academy. Repéré à:
https://www.heacademy.ac.uk/system/files/resources/id477_aligning_teaching_for_constructive_learning.pdf

Biggs, J. (2003, avril). *Aligning teaching and assessing to course objectives*. Actes du colloque “Teaching and Learning in Higher Education: New Trends and Innovations”. University of Aveiro: Portugal. Repéré à:
<https://pdfs.semanticscholar.org/edd0/9f550a0cef684a4ea0bfeefdcc25cc43d90d.pdf>

Biggs, J. & Tang, C. (2007). *Teaching for quality learning at university: what the student does*. New York, NY: McGraw-Hill.

Biggs, J. & Tang, C. (2009). *Applying constructive alignment to outcomes-based teaching and learning*. Repéré à :
https://intranet.tudelft.nl/fileadmin/Files/medewerkersportal/TBM/Onderwijsdag_2014/What-is-ConstructiveAlignment.pdf.

Biggs, J. & Tang, C. (2011). *Teaching for Quality Learning at University*. New York: McGraw Hill.

Black, P., Harrisson, C., Hodgen, J., Marshall, B. & Serret, N. (2010). Validity in teachers’ summative assessments. *Assessment in Education Principles, Policy & Practice*, 17 (2), 215-232.

Blanchet, A. (2010). La réforme à l’épreuve de l’évaluation dans le canton de Vaud. In P. Gilliéron Giroud & L. Ntamakiliro (Ed.), *Réformer l’évaluation scolaire : mission impossible?* (pp. 79-111). Berne : Peter Lang.

Bloom, B. J. (1969). *Taxonomie des objectifs pédagogiques*, vol.1. Domaine cognitif. Montréal : Les Presses de l’université de Québec.

Bloom, B. J. (1971). *Handbook on formative and summative evaluation of student learning*. New York : McGraw-Hill Book Co.

Bloxham, S., Boyd, P. & Orr, S. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36 (6), 655-670.

Bodin, A. (2010). *Taxonomie pour les énoncés de mathématiques - Classement par niveaux hiérarchisés de complexité cognitive*. Repéré à : http://www.irem.univ-mrs.fr/IMG/pdf/Taxonomie_A-_Bodin2010.pdf.

Bonner, S. M. (2013). Validity in Classroom Assessment: Purposes, Properties, and Principles. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 87-106). Thousand Oaks, CA: SAGE Publications.

Boucenna, S. & Charlier, E. (2012). Accompagnement d'équipes d'enseignants, entre recherche et développement professionnel. In E. Charlier & S. Biemar (Eds.), *Accompagner : un agir professionnel* (pp. 127-138). Bruxelles : De Boeck.

Bourassa, B. Leclerc, Ch. & Fournier, G. (2012). Assumer et risquer une posture de recherche collaborative. In B. Bourassa & M. Boudjaoui (Ed.), *Des recherches collaboratives en sciences humaines et sociales : enjeux, modalités et limites* (pp. 13-46). Québec: Presses Universitaires de Laval.

Boutin, G. (2006). *L'entretien de recherche qualitatif*. Sainte-Foy : Québec : Presses de l'Université du Québec.

Braxmeyer, N., Guillaume, J.-C. & Levy, J.-F. (2005). Les pratiques d'évaluation des enseignants au collège. *Les dossiers du Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, Direction de l'évaluation et de la prospective*, 160, 5-117.

Brookhart, S. M. (2002). What will teachers know about assessment, and how will that improve instruction. In R. W. Lizzitz & W. D. Schafer (Eds.), *Assessment in educational reform: Both means and ends* (pp. 2-17). Boston, MA: Allyn & Bacon.

Brookhart, S. M. (2003). Developing Measurement Theory for Classroom Assessment Purposes and Uses. *Educational Measurement : Issues and Practice*, 22 (4), 5-12.

Brookhart, S. M. (2005). Assessment Theory for College Classrooms. *New directions for*

teaching and learning, 100, 5-14.

Brookhart, S. M. (2011). Starting the conversation About Grading. *Effective Grading Practices*, 69 (3), 10-14.

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement : Issues and Practice*, 30 (1), 3-12.

Brookhart, S. M. (2013). Grading. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257-271). Thousand Oaks, CA: SAGE Publications.

Brookhart, S., Guskey, T. R., Bowers, A. J. McMillan, J. H., Smith, J. K. & Smith, (...). (2016). A Century of Grading Research : Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86 (4), 803-848.

Brookhart, S. (2017). *How to use grading to improve learning*. Virginia : ASCD.

Brophy, J. E. (2000). *Teaching, Educational Practices Series, 1*, Repéré à <http://files.eric.ed.gov/fulltext/ED440066.pdf>

Butera, F., Buchs, C. & Darnon, C. (2011). *L'évaluation, une menace ?* Paris : PUF.

Buyse, A. & Vanhulle, S. (2009). Ecriture réflexive et développement professionnel : quels indicateurs ? *Questions vives*, 5, 11, 225-242.

Campbell, C. (2013). Research on Teacher Competency in Classroom Assessment. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 71-84). Thousand Oaks, CA: SAGE Publications.

Cardinet, J. (1984). Le découpage des objectifs pédagogiques. *Mesure et évaluation en éducation*, 7 (2), 5-48.

Cardinet, J. (1989). Evaluer sans juger. *Revue française de pédagogie*, 88, 41-52.

Cardinet, J. (1990). *Remettre le quantitatif à sa place en évaluation scolaire*. Neuchâtel : IRDP.

Cardinet, J. & Laveault, D. (2001). L'activité évaluative en éducation : évolutions des

préoccupations des deux côtés de l'Atlantique. In G. Figari & M. Achouche (Ed.), *L'activité évaluative réinterrogée* (pp. 15-29). Bruxelles : De Boeck.

Carette, V. (2007). Les implications de la notion de compétence sur l'évaluation. *Education-Formation*, 286, 51-61.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. London : Routledge.

Carulla, C., Corti, D., de Pietro, J.-F., Kassam, S., Roth, M., Sanchez Abchi, V. & Singh, L. (2013). *Quoi et comment évaluer en référence au Plan d'études romand (PER) : premières pistes* (Rapport scientifique intermédiaire). Neuchâtel : IRDP.

Charlier, E. (2012). Former des enseignants professionnels pour une formation continuée articulée à la pratique. In L. Paquay, M. Altet, E. Charlier & Ph. Perrenoud (Eds.), *Former des enseignants professionnels, quelles stratégies ? Quelles compétences ?* (pp. 119-143). Bruxelles : De Boeck.

Charlier, E. & Biemar, S. (2012). *Accompagner un agir professionnel*. Bruxelles : De Boeck.

Chevallard, Y. (1989). Evaluation, véridiction, objectivation. In J. Colomb & J. Marsenach (Eds.), *L'évaluateur en révolution* (pp.13-36). INRP : Paris.

Christie, C. A. (2003). What Guides Evaluation ? A Study of How Evaluation Practice Maps onto Evaluation Theory. *New directions for evaluation*, 97, 7-34.

Cifali, M. & Myftiu, B. (2006). *Dialogues. Récits d'éducation sur la différence*. Nice : Éditions Les paradigmes.

Cohen, S. A. (1987). Instructional Alignment: Searching for a Magic Bullet. *Education Researcher*, 16 (8), 16-20.

Cohen, S. A. (1995). Instructional Alignment. In J. H. Block, S. T. Stevenson & T. R. Guskey, *School Improvement Programs. A Handbook for Educational Leaders* (pp. 153-181). New York: Scholarship Leadership Policy Research.

Coles, C. (2002). Developing professional judgment. *The Journal of Continuing Education in*

the Health Professions, 22, 3-10.

Conseil national d'évaluation du système scolaire (CNESCO). (2014). *L'évaluation des élèves par les enseignants dans la classe et les établissements : réglementation et pratiques. Une comparaison internationale dans les pays de l'OCDE*. Paris: CNESCO.

Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.

Crahay, M. (2006). En quoi les curriculums peuvent-ils être objets d'investigation scientifique ? In F. Audigier, M. Crahay & J. Dolz (Eds.), *Curriculum, enseignement et pilotage* (pp. 7-37). Bruxelles : De Boeck.

Crahay, M. (2010). Les réformes pédagogiques échouent et pourtant l'école change ! In P. Gilliéron Giroud & L. Ntamakiliro (Ed.), *Réformer l'évaluation scolaire : mission impossible ?* (pp. 231-264). Berne : Peter Lang.

Crinon, J. & Guigue, M. (2006). Ecriture et professionnalisation. *Revue française de pédagogie*, 156, 117-169.

Cros, F. & Raisky, C. (2010). Autour des mots de la formation « Référentiel ». *Recherche & Formation*, 64, 105-116.

Cross, L. H. (1995). Grading students. *Practical Assessment, Research & Evaluation*, 4, 1-4.

Cross, L. H. & Frary, R. B. (1999). Hodgepodge Grading : Endorsed by Students and Teachers Alike. *Applied Measurement in Education*, 12 (1), 53-72.

Crotty, M. (1998). *The foundations of social research*. Thousand Oaks : Sage.

Cuche, C., Canivet, C. & Donnay, J. (2012). L'accompagnement, source de création de savoirs. In E. Charlier & S. Biemar (2012). *Accompagner un agir professionnel* (pp. 139-152). Bruxelles : De Boeck.

Denzin, N. K. (1978). *The research act : a theoretical introduction to sociological methods*. New York : Mc Graw-Hill.

Dechamboux, L. & Mottier Lopez, L. (2015). L'enquête de l'enseignant-évaluateur et du chercheur : l'abduction au cœur du processus et les raisons de les comparer. *Recherches qualitatives*, 34 (1), 173-195.

De Ketele, J.-M. (2012). A la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ? In L. Mottier Lopez & G. Figari (Eds.). *Modélisations de l'évaluation en éducation* (pp. 195-210). Bruxelles : De Boeck.

De Ketele, J.-M. & Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et Evaluation en Education*, 28 (3), 1-26.

De Ketele, J.-M. & Roegiers, X. (2009). *Méthodes du recueil d'informations*. Bruxelles : De Boeck.

De Landsheere, G. (1992). *Evaluation continue et examens, précis de docimologie*. Bruxelles : Labor.

DeLuca, C. & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning, *Assessment in Education: Principles, Policy & Practice*, 17 (4), 419-438.

DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2015). Teacher assessment literacy: a review of international standards and measures, *Educational Assessment, Evaluation and Accountability*, 28 (3), 251-272.

DeLuca, C. & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability, *Assessment in Education: Principles, Policy & Practice*, 24 (2), 121-126.

Demeuse, M. & Strauven, S. (2006). *Développer un curriculum d'enseignement ou de formation: Des options politiques au pilotage*. Bruxelles : De Boeck.

Demonty, I., Fagnant, A. & Dupont, V. (2015). Analyse d'un outil d'évaluation en mathématiques : entre une logique de compétences et une logique de contenu. *Mesure et évaluation en éducation*, 38, (2), 1-29.

Denner, P. R., Salzman, S. A., & Harris, L. B. (2002, février). *Teacher work sample assessment: An accountability method that moves beyond teacher testing to the impact of*

teacher performance on student learning. Communication présentée à “the annual meeting of the American Association of Colleges for Teacher Education”, New York.

Desbiens, J.-F., Borges, C. et Spallanzani, C. (Dir.) (2012). *J'ai mal à mon stage. Problèmes et enjeux de la formation pratique en enseignement. 21 cas étudiés*. Québec : Presses de l'Université du Québec.

Desclaux, B. (2014). Esprit d'entreprise et défiance éducative. *Administration et éducation*, 141, 23-29.

Desgagné, S. (1997). Le concept de recherche collaborative : l'idée d'un rapprochement entre chercheurs universitaires et praticiens enseignants. *Revue des sciences de l'éducation*, 23 (2), 371-393.

Desgagné, S. (2005). *Récits exemplaires de pratique enseignante : analyse typologique*. Sainte-Foy : Presses de l'université du Québec.

Desgagné, S. (2007). Le défi de coproduction de « savoir » en recherche collaborative : autour d'une démarche de reconstruction et d'analyse de récits de pratique enseignante. In M. Anadon (Ed.), *La recherche participative : multiples regards* (pp. 89-121). Québec : Les Presses de l'Université du Québec.

Desgagné, S., Bednarz, N., Lebus, P., Poirier, L. & Couture, C. (2001). L'approche collaborative de recherche en éducation : un rapport nouveau à établir entre recherche et formation. *Revue des sciences de l'éducation*, 27 (1), 33-64.

Desgagné, S., Gervais, F. et Larouche, H. (2001). L'utilisation du “récit” de “pratique” : son potentiel pour le développement professionnel des enseignants et autres éducateurs du monde scolaire. In A. Beauchesne, S. Martineau et M. Tardif (dir.), *La recherche en éducation et le développement de la pratique professionnelle en enseignement* (pp. 203-223). Sherbrooke : Éditions du CRP.

Desgagné, S. & Larouche, H. (2010). Quand la collaboration de recherche sert la légitimation d'un savoir d'expérience. *Recherches en Education*, 1, 7-18.

De Vecchi, G. (2011). *Evaluer sans dévaluer*. Paris : Hachette Education.

- Donnay, J. & Charlier, E. (2008). *Apprendre par l'analyse de pratiques : Initiation au compagnonnage réflexif*. Namur : Presses universitaires de Namur.
- Dubet, F. (1994). *Sociologie de l'expérience*. Paris : Seuil.
- Dubois, N. & Marceau, R. (2005). Un état des lieux théoriques de l'évaluation : une discipline à la remorque d'une révolution scientifique qui n'en finit pas. *The Canadian Journal of Program Evaluation*, 20 (1), 1-36.
- Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Paris : Armand Colin.
- Durand, M.-J. & Chouinard, R. (2012). *L'évaluation des apprentissages*. Québec : HMH.
- Duru-Bellat, M. (2010). Ce que la mixité fait aux élèves. *Revue de l'OFCE*, 3 (114), 197-212.
- Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society*, 51, 599-635.
- Edmonds, R. R. (1979). Effective schools for urban poor. *Educational Leadership*, 37 (10), 15-24.
- Edwards, F. (2017). A rubric to track the development of secondary preservice and novice teachers' summative assessment literacy, *Assessment in Education: Principles, Policy & Practice*, 24 (2), 205-227.
- Figari (2006). Les référentiels entre théorie et méthodologie. In G. Figari & L. Mottier Lopez (Eds.), *Recherche sur l'évaluation en éducation* (pp. 101-116). Paris : L'Harmattan.
- Figari, G. (2012). Configurations pour une épistémologie de l'évaluation. In L. Mottier Lopez & G. Figari (Eds.), *Modélisations de l'évaluation en éducation* (pp. 211-224). Bruxelles : De Boeck.
- Figari, G. (2013). L'évaluation entre « technicité » et « théorisation ». *Mesure et évaluation en éducation*, 36 (3), 5-24.
- Figari, G. & Remaud, D. (2014). *Méthodologie d'évaluation en éducation et formation*.

Bruxelles : De Boeck.

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural-historical perspective. *Assessment in Education: Principles, Policy & Practice*, 22 (2), 224-246.

Flick, U. (1998). *An introduction to qualitative research*. London : Sage.

Francis, V. (2011, octobre). Le récit d'expérience en formation des enseignants : usages et apports en vue d'une articulation des savoirs issus de l'expérience et des savoirs de la recherche. Communication présentée au colloque de l'AECSE « Crise et/en éducation. Épreuves, controverses et enjeux nouveaux », Paris : Repéré à : <http://colloque-crise-aecse-2011.eu/communications/liste>.

Fugier, P. (2010). Les approches compréhensives et cliniques des entretiens sociologiques. *Revue Interrogations*, 11. Repéré à : <http://www.revue-interrogations.org/Les-approches-comprehensives-et>.

Fullan, M. (2001). *The new meaning of educational change*. New York, NY: Teachers College Press.

Gadamer, H. G. (1996). *Vérité et méthode. Les grandes lignes d'une herméneutique philosophique*. Paris : Seuil.

Gagné, Ph., Dumont, L., Brunet, S. & Boucher, G. (2013). Curriculum Alignment: Establishing Coherence. *Collected Essays on Learning and Teaching*, VI, 7-12.

Gagnon, R., Monnier, A. & Dolz, J. (2014). L'évaluation dans la formation à l'enseignement de la production écrite : l'exemple suisse romand », *Pratiques* [En ligne], 161-162 | 2014, mis en ligne le 05 juin 2014, consulté le 04 mars 2017. URL : <http://pratiques.revues.org/2120> ; DOI : 10.4000/pratiques.2120

Gauthier, C., Mellouki, M., Bissonnette, S. & Richard, M. (2005). *Ecoles efficaces et réussite scolaire des élèves à risque. Un état de la recherche*. (Rapport de recherche du Centre de recherche interuniversitaire sur la formation et la profession enseignante). Québec : CRIFPE, Université Laval.

Geminard, L. & Martinand, J.-L. (1987). Connaître et transformer la matière. *Revue française de pédagogie*, 81, 113-115.

Gerard, F.-M. (2009). *Evaluer des compétences*. Bruxelles : de Boeck.

Gervais, F. et Desgagné, S. (2003). The reconstruction process of teachers' narratives. *Policy and Practice in Education*, 10 (1), 45-59.

Gil, F. (1998). La bonne description. *Enquête*, 6, 129-152.

Gimonnet, B. (2007). *Les notes à l'école ou le rapport à la notation des enseignants de l'école élémentaire*. Paris : L'Harmattan.

Goasdoué, R., Vantourout, M. & Bedoin, D. (2016). La correction de dissertations : nouveau regard sur la construction des jugements des évaluateurs. In L. Mottier Lopez & W. Tessaro (Ed.), *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages* (pp. 73-93). Berne : Peter Lang.

Gohier, Ch. (2004). De la démarcation entre critères d'ordre scientifique et d'ordre éthique en recherche interprétative. *Recherches qualitatives*, 24, 3-17.

Goldberg, G. L. & Roswell, B. S. (2000). From perception to practice : The impact of Teacher's scoring experience on performance-based instruction and classroom assessment. *Educational Assessment*, 6, 257-290.

Gomoran, A., Porter, A.C, Smithson, J, & White, P.A (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis*, 19, 325-338.

Good, R. (2011). Formative Use of Assessment Information: It's a Process, So Let's Say What We Mean. *Practical Assessment, Research & Evaluation*, 16 (3), 1-6.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33 (2), 14-18.

Graue, M. E. & Walsh, D. J. (1998). *Studying children in context. Theories, methods and ethics*. London : Sage.

Grisay, A. (1984). Les mirages de l'évaluation scolaire (1). Rendement en français, notes et échecs à l'école primaire ? *Revue de la Direction générale de l'Organisation des Etudes*, XIX (5), 29-42.

Grisay, A. (1988c). *Du mythe de la « bonne école » à la réalité (fuyante) de l'« école efficace »*. Liège : Service de Pédagogie expérimentale de l'Université.

Grisay, A. (1989). Améliorer l'évaluation-bilan à l'école primaire. *Mesure et évaluation en éducation*, 11 (4), 5-24.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Thousand Oaks : Sage Publications.

Guichon, N. (2007). *Récits de soi en formation – L'écriture réflexive dans la formation didactique des enseignants*. Actes du 27^e colloque « La biographie, la réflexivité et les temporalités : articuler langues, culture et formation. Tours : France. Repéré à : <http://pagesperso-orange.fr/nicolas.guichon/topic4/guichon.2007.r-flexivit-.pdf>.

Guskey, T.R. (2000). *Evaluating Professional Development*. Thousand Oaks: Corwin Press.

Guskey, T. R. (2003). How Classroom Assessments Improve Learning. *Educational Leadership*, 60 (5), 7-11.

Guskey, T. R. (2006). Making High School Grades Meaningful. *Phi Delta Kappan*, 87 (9), 670-675.

Guskey, T. R. (2009, avril). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Communication présentée à l'American Educational Research Association, San Francisco.

Guskey, T. R. (2011). Five Obstacles to Grading Reform. *Effective Grading Practices*, 69 (3), 16-21.

Guskey, T. R., Swan, G. M. & Ann Jung, L. (2011). Grades that mean something. *Phi Delta Kappan*, 93 (2), 53-57.

Hadji, Ch. (1997). *L'évaluation démystifiée*. Paris : ESF.

Hadji, Ch. (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles : De Boeck.

Hadji, C. (2016). Le système éducatif peut-il passer de l'évaluation normative à une évaluation constructive ? In B. Bocquet (Ed.), *La fièvre de l'évaluation quels symptômes, quels traitements ?* (pp.113-135). Villeneuve d'As : Presses Universitaires du Septentrion.

Hadji, C. (2017). Savoir mettre en œuvre une évaluation constructive. In A. Bentolila (Ed.), *L'essentiel de la pédagogie*. (pp. 229-250). Paris : Nathan.

Hammerness, K. (2006). From Coherence in Theory to Coherence in Practice. *Teachers College Record*, 108 (7), 1241-1265.

Harlen, W. (2005). Teacher's summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16 (2), 207-223.

Harlen, W. (2012). On the relationship between assessment or formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 87–102). London: Sage.

Hermann, J. L., Webb, N. & Zuniga, S. A. (2007). Measurements issues in the alignment of standards and assessments : A case study. *Applied Measurement in Education*, 20, 101-126.

Jaworski B. (2003). Research practice into/influencing mathematics teaching and learning development: towards a theoretical framework based on co-learning partnerships. *Educational Studies in Mathematics*, 54 (2-3), 249-282.

Jeffrey, D. (2013). *L'éthique dans l'évaluation scolaire*. Québec : Les Presses de l'Université Laval.

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment, *Assessment & Evaluation in Higher Education*, 39 (7), 840-852.

Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences, *Educational Research Review*, (2), 130–144.

Jorro, A. (2002). Ecrire en formation. *Les Cahiers de Pédagogie Expérimentale*, 11-12, 11-31.

- Jorro, A. (2013). L'accompagnement des enseignants dans l'activité évaluative face à des situations de production écrite. *Revue française de linguistique appliquée*, (XVIII), 107-116.
- Jorro, A. (2015, 4 septembre). *Des compétences aux gestes évaluatifs des enseignants*. [vidéo en ligne]. Repéré à <https://www.hepl.ch/cms/accueil/actualites-et-agenda/actu-hep/les-conferences-du-colloque-geva.html>.
- Jorro, A. (2016). Se former à l'activité évaluative. *Education permanente*, 208, 53-64.
- Kahn, S. (2010). Les réformes contre le redoublement ou les limites des recherches non collaboratives. *Recherches en Education, Hors Série 1*, 71-84
- Karsenti, T. (2006). Pragmatisme et méthodologie de recherche en sciences de l'éducation : passons à la version 3.0. *Formation et profession*, 13 (1), 2-5.
- Kaufmann, J.-C. (2011). *L'entretien compréhensif*. Paris : Armand Colin.
- Kaufmann, J.-C. (2008). *Quand Je est un autre. Pourquoi et comment ça change en nous*. Paris : Armand Colin.
- Kilday, C. R., Kinzie, M. B., Mashburn, A. J. & Whittaker, J. V. (2011). Accuracy of teacher judgements of preschoolers' math skills. *Journal of Psychoeducational Assessment*, 29 (4), 1-12.
- Kitabgi, S. (2009). *L'évaluation scolaire est-elle au service de l'orientation ?* Paris : Centre d'orientation de CCIP.
- Klapp Lekholm, A. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research*, 55 (6), 587-608.
- Klein, S. P., Stecher, B. M., Shavelson, R. J., Mc Caffrey, D., Ormseth, T., Bell, R. M., Comfort, K. & Othman, A. R. (1998). Analytic Versus Holistic Scoring of Science Performance Tasks. *Applied Measurement in Education*, 11 (2), 121-137.
- Klenowski, V. & Wyatt-Smith, C. (2014). *Assessment for Education*. Thousand Oaks, CA: SAGE Publications.

- Klinger, D. A., Volante, L. & DeLuca, C. (2009). *Impact evaluation of the classroom assessment workshop series*. Toronto : Elementary Teachers' Federation of Ontario.
- Knight, P. & Yorke, M. (2003). *Assessment, Learning and Employability*. Maidenhead : Open University Press.
- Korthagen, F. A. J. (2001, avril). *Linking practice and theory : the pedagogy of realistic teacher education*. Communication présentée à l'Annual Meeting of the American Educational Research Association, Seattle.
- Kuhn, T. S. (1970/2008). *La structure des révolutions scientifiques*. Paris : Champs Flammarion.
- Lafontaine, D. (2001). Quoi de neuf en littérature ? Regard sur trente ans d'évaluation de la lecture. *Les cahiers du Service de pédagogie expérimentale*, 7 (8), 71-95.
- Lafortune L. & Allal L. (dir.) (2008). *Jugement professionnel en évaluation : Pratiques enseignants au Québec et à Genève*. Sainte-Foy : Presses de l'université du Québec.
- Lafortune, L. & Turcotte, S. (2008). *Accompagner l'évaluation des apprentissages dans l'école québécoise : accompagnement d'équipes-cycle pour des pratiques évaluatives de développement des compétences*. Repéré sur le site de l'UQTR : https://oraprdnt.uqtr.quebec.ca/pls/public/docs/GSC1/F194971554_fasc5eval_eq_cycle_oct8.pdf.
- Laveault, D. (2004, juin). *Evaluer les apprentissages : un jeu de serpents et échelles ?* Communication présentée au 24^e colloque de l'AQPC, Montréal. Repéré à : http://www.cdc.qc.ca/actes_aqpc/2004/laveault_dany_900.pdf.
- Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30 (3), 483-500.
- Laveault, D. & Allal, L. (2016). Implementing Assessment for Learning : Theoretical and Practical Issues. In D. Laveault & L. Allal (Ed.). *Assessment for Learning : Meeting the Challenge of Implementation*. (pp. 1-18). New York : Springer.

Laveault, D. & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles : De Boeck.

Laveault, D., & Miles, C. (2008). Utilité des échelles descriptives et différences individuelles dans l'autoévaluation de l'écrit. *Mesure et évaluation en éducation*, 31 (1), 1-29.

Laveault, D. (2010). L'évaluation en salle de classe : des politiques aux pratiques. *Mesure et évaluation en éducation*, 32 (3), 150 pages.

Le Boterf, G. (2001). *Construire les compétences individuelles et collectives*. Paris : Éditions d'Organisation.

Létor, C. & Périsset Bagnoud, D. (2010). Travailler ensemble entre logiques professionnelles, organisationnelles et institutionnelles : un développement professionnel sous contraintes. In L. Corriveau, C. Létor, D. Périsset Bagnoud & L. Savoie-Zajc (Dir.), *Travailler ensemble dans les établissements scolaires et de formation* (pp. 165-173). Bruxelles : De Boeck.

Legendre, M.-F. (2008). La notion de compétence au cœur des réformes curriculaires : effet de mode ou moteur de changements en profondeur ? In F. Audigier et al., *Compétences et contenus* (pp. 27-50). Bruxelles : De Boeck.

Lenoir, Y. (2004). L'enseignant expert. Regard critique sur une notion non dépourvue d'intérêt pour la recherche sur les pratiques enseignantes. *Recherche et formation*, 47, 9-23.

Lenoir, Y. (2011). Conséquences des conceptions curriculaires actuelles sur les modes évaluatifs. *Les Dossiers des Sciences de l'Éducation*, 25, 13-28.

Lessard C. (1999, octobre). *Professionnalisation, déprofessionnalisation, reprofessionnalisation, déqualification, prolétarianisation, nouvelle régulation de l'éducation : que se passe-t-il au juste ?* Communication prononcée au séminaire Girsef, Louvain-la-Neuve.

Lessard, C. & Tardif, M. (2004). Les transformations actuelles de l'enseignement : trois scénarios possibles dans l'évolution de la profession enseignante. In M. Tardif, & C. Lessard, *La profession d'enseignant aujourd'hui. Evolutions, perspectives et enjeux internationaux* (pp. 265-288). Laval : Presses de l'Université Laval.

- Li, L., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap : The challenges of aligning science instruction with standards and tests. *Science Educator*, 15 (1), 1-12.
- Lieberman, A. (1986). Collaborative research : Working with, not working on... *Educational Leadership*, 43 (5), 29-32.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills: SAGE publication.
- Linn, R. & Miller, M. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23 (2), 26-32.
- Machado, E. A. (2007). *Evaluation et participation. Une étude sur le rôle des acteurs dans l'évaluation de la formation continue*. Thèse de doctorat en Sciences de l'éducation, Universidade do Minho.
- Malet, R. (2015). Politiques d'accountability et devenir de la profession enseignante aux Etats-Unis. *Evaluer. Journal international de Recherche en Education et Formation*, 1 (1), 27-44.
- Marc, V. & Wirthner, M. (2013). *Développement d'un modèle d'évaluation adapté au PER*. (Rapport scientifique intermédiaire). Neuchâtel : IRDP.
- Marcel, J.-F. (2011). Des interrelations entre les processus curriculaires et les processus évaluatifs. *Les Dossiers des Sciences de l'Education*, 25, 7-12.
- Marcel, J.-F. (2014). Les pratiques enseignantes mises en récit. *Recherches en éducation*, 19, 82-95.
- Maroy C. & Cattonar B. (2002). Professionnalisation ou déprofessionnalisation des enseignants ? Le cas de la Communauté française de Belgique. *Cahier de recherche en Education et Formation*, 18, 4-26.

- Martinand, J.-L. (1986). *Connaître et transformer la matière*. Berne : Peter Lang.
- Martinez, J. F., Stecher, B. & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence in the ECLS. *Educational Assessment, 14*, 78-102.
- Marton, F., & Tsui, A. B. (2004). *Classroom discourse and the space of learning*. London: Lawrence Erlbaum Associates.
- Martone, A. & Sireci, S. G. (2009). Evaluating Alignment Between Curriculum, Assessment, and Instruction, *Review of Educational Research, 79* (4), 1332-1361.
- Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks : Corwin Press.
- Marzano, R. J. (2002). A comparison of Selected Methods of Scoring Classroom Assessments. *Applied Measurement in Education, 15* (3), 249-267.
- Marzano, R. J. & Heflebower, T. (2011). Grades That Show What Students Know, *Effective Grading Practices, 69* (3), 34-39.
- Maulini, O. (2012). Resserrer ou justifier les classements ? L'évaluation scolaire entre deux injonctions. *Diversité, 169*, 133-137.
- Mc Elligott, J. & Brookhart, S. M. (2009). Legal issues of grading in the era of high stakes accountability. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 57-74). Thousand Oaks, CA : Corwin Press.
- Mc Gatha, M. B. & Bush, S. (2013). Classroom Assessment in Mathematics. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 449-460). Thousand Oaks, CA: SAGE Publications.
- Mc Kinney, S. E., Chappell, S., Berry, R. Q. & Hickmann, B. T. (2009). An examination of the instructional practices of mathematics teachers in urban schools. *Preventing School Failure: Alternative Education for Children and Youth, 53* (4), 278-284.
- McMillan, J. H. (2003). *The relationship between instructional and classroom assessment*

practices of elementary teachers and students scores on high-stakes tests (Report ERIC Document Reproduction Service No. ED472164). Repéré à <https://files.eric.ed.gov/fulltext/ED472164.pdf>

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22 (4), 34–43.

McMillan, J. H. (2005). *The impact of high-stakes test results on teachers' instructional and classroom practices* (Report ERIC Document Reproduction Service No. ED490648). Repéré à : <https://files.eric.ed.gov/fulltext/ED490648.pdf>

Mc Millan, J. H. (2013). Why we need research on classroom summative assessment. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 3-16). Thousand Oaks, CA: SAGE Publications.

McMillan, J. H. & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making* (Report ERIC Document Reproduction Service No. ED447195). Repéré à : <https://files.eric.ed.gov/fulltext/ED447195.pdf>

Meier, S. L., Rich, B. S., Cady, J. (2006). Teacher's use of rubrics to score non-traditional tasks : factors related to discrepancies in scoring. *Assessment in Education*, 13 (1), 69-95.

Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : PUF.

Merle, P. (2007). *Les notes. Secrets de fabrication*. Paris : PUF.

Merle, P. (2012a). L'évaluation par les notes : quelle fiabilité et quelles réformes ? *Regards croisés sur l'économie*, 2 (12), 218-230.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference ? *American Secondary Education*, 33 (1), 49-64.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12 (2), 101-113.

Mertler, C. A., & Campbell, C. (2005, avril). *Measuring teachers' knowledge and application*

of classroom assessment concepts: Development of the “Assessment Literacy Inventory.”

Communication présentée à the Annual Meeting of the American Educational Research Association, Montreal, Québec.

Millon-Fauré, K. (2013). Processus de négociation didactique et mesure du niveau des élèves : des fonctions concurrentes de l'évaluation, *Carrefours de l'éducation*, 36 (2), 149-166.

Mitchell, S. N., Reilly, R. C. & Logue, M. E. (2009). Benefits of collaborative action research for the beginning teacher. *Teaching and Teacher Education*, 25, 244-349.

Morissette, D. (1997). *Guide pratique de l'évaluation sommative*. Bruxelles : De Boeck.

Morissette, J. & Desgagné, S. (2009). Les jeux des positions de savoir en recherche collaborative : une analyse des points de vue négociés d'un groupe d'enseignantes du primaire. *Recherches qualitatives*, 28 (2), 118-144.

Morissette, J. & Tessaro, W. (2015). Apports des recherches collaboratives sur le développement professionnel des enseignants et des formateurs en matière d'évaluation. *Evaluer. Journal international de Recherche en Education et Formation*, 1 (2), 9-14.

Moskal, B. M. & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*, 7 (10), 71-81.

Moss, C. (2003). Reconceptualizing Validity for Classroom Assessment. *Educational Measurement: Issues and Practice*, 22 (4), 13-25.

Moss, C. (2013). Research on classroom summative assessment. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 235-256). Thousand Oaks, CA: SAGE Publications.

Mottier Lopez, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30 (3), 465-482.

Mottier Lopez, L. (2009). L'évaluation en éducation : des tensions aux controverses. In L. Mottier Lopez & M. Crahay (Ed.), *Evaluations en tension* (pp. 7-25). Bruxelles : De Boeck.

Mottier Lopez, L. (2010). Perspectives conclusives. Repenser la complémentarité des

évaluations formative et certificative des compétences professionnelles. In L. Bélair, C. Lebel, N. Sorin & A. Roy (Ed.), *Régulation et évaluation des compétences en enseignement* (pp. 237-249). Ottawa : Presses Universitaires d'Ottawa.

Mottier Lopez, L. (2013). Pour une culture partagée de l'évaluation. Du retour à des pratiques d'évaluation traditionnelles vers la construction d'un nouveau rapport entre la recherche, le terrain et le politique. *Revue Education Canada*. En ligne <http://www.cea-ace.ca/education-canada/article/pour-une-culture-partagée-de-l'évaluation>.

Mottier Lopez, L. (2013). De la mesure à l'évaluation collaborative en éducation. *Revue française d'administration publique*, 148 (4), 939-952.

Mottier Lopez, L. (2014). L'évaluation pédagogique va-t-elle enfin marcher sur ses deux pieds ? Les enseignements de l'histoire récente de l'école genevoise. *Education et francophonie*, XLII, (3), 85-101.

Mottier Lopez, L. (2015). *Evaluations formative et certificative des apprentissages, enjeux pour l'enseignement*. Bruxelles : De Boeck.

Mottier Lopez, L. (2015). Tâches pour apprendre, tâches pour évaluer. *Revue des Hautes écoles pédagogiques et institutions assimilées de Suisse romande et du Tessin*, 19, 55-65.

Mottier Lopez, L. (2016). Quels référentiels pour évaluer les élèves au plus près de ce qu'ils savent ? *irdp FOCUS*, 1-4. Repéré à : <https://www.irdp.ch/institut/quels-referentiels-pour-evaluer-eleves-plus-pres-1916.html>.

Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30 (3), 465-482.

Mottier Lopez, L. & Allal, L. (2010). Le jugement professionnel en évaluation : quelles triangulations méthodologiques et théoriques ? In L. Paquay, C. Van Nieuwenhoven & P. Wouters (Ed.), *L'évaluation, levier du développement professionnel ? Tensions, dispositifs, perspectives* (pp. 237-250). Bruxelles : De Boeck.

Mottier Lopez, L. & Dechamboux, L. (2017). D'un référentiel d'évaluation fixe à une co-

constitution référentielle dynamique, ce que nous apprend le jugement situé de l'enseignant. *Contextes et Didactiques*, 9, 12-29.

Mottier Lopez, L. & Figari, G. (Eds.) (2012). *Modélisations de l'évaluation en éducation*. Bruxelles : de Boeck.

Mottier Lopez, L. & Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et Evaluation en éducation*, 31 (3), 5-34.

Mottier Lopez, L. & Morales Villabona, F. (2016). Teachers' professional development in the context of collaborative research : Toward practices of collaborative assessment for learning in the classroom. In D. Laveault & L. Allal (Ed.), *Assessment for learning in the classroom : Meeting the challenge of implementation* (pp. 161.180). London : Springer.

Mottier Lopez, L. & Pasquini, R. (2017). Professionnal controversies between teachers about their sommative assessment practices. *Assessment in Education: Principles, Policy & Practice*, 24 (2), 228-249.

Mottier Lopez, L., Serry, S. & Sales Cordeiro, G. (2016). Evaluation certificative de productions écrites d'élèves : référence au modèle didactique du genre textuel dans l'épreuve et lors de la correction de faits divers. In E. Falardeau, J. Dolz, J.-L. Dumortier & P. Lefrançois, *L'évaluation en classe de français, outil didactique et politique* (pp. 27-61). Namur : Presses Universitaires de Namur.

Mottier Lopez, L., Tessaro, W., Dechamboux, L. & Morales Villabona, F. (2012). La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves. *Questions vives*, 18, 159-175.

Mottier Lopez, L. & Tessaro, W. (Eds.).(2016). *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages*. Berne : Peter Lang.

Mucchielli, A. (2009). *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*. Paris : Armand Colin.

Muller, A. (2006). Définir le curriculum : une approche pragmatiste. In F. Audigier, M.

Crahay & J. Dolz (Ed.), *Curriculum, enseignement et pilotage* (pp. 99-115). Bruxelles : De Boeck.

Neumeyer, M. & Vellas, E. (2015). Dé-chiffrer l'humain. In. M. Neumeyer & E. Vellas (Eds.), *Evaluer sans noter, éduquer sans exclure* (pp. 7-20). Lyon : Chronique Sociale.

Nizet, I. & Leroux, J.-L. (2015). La construction de savoirs conceptuels en évaluation : enjeux pour un développement professionnel en contexte de formation continue. *Evaluer. Journal international de Recherche en Education et Formation*, 1 (2), 15-29.

Noizet G. (1961). Etude docimologique sur la correction de l'écrit du baccalauréat. *Bulletin de l'Institut national d'étude du travail et d'orientation professionnelle*, 4, 257-267.

Noizet, G. & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.

Nusche, D., Radinger, T., Santiago, P. & Shewbridge C. (2013). *Synergies for better learning. An international perspective on evaluation and assessment*. OECD Reviews of Evaluation and Assessment in Education. Paris : OCDE.

OCDE (2000). *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*. Repéré à : <http://www.oecd.org/fr/edu/innovation-education/39438013.pdf>.

Ohlsen, M. T. (2007). Classroom assessment practices of secondary school members of NCTM. *American Secondary Education*, 36 (1), 4-14.

Orellana, I. (2002). *La communauté d'apprentissage en éducation relative à l'environnement : signification, dynamique, enjeux* (Thèse de doctorat inédite en sciences de l'éducation). Université du Québec à Montréal.

Osborn, M. (2002). Le métier d'enseignant en période de réforme : le cas anglais. *Revue Internationale d'Education*, 30, 27-39.

Osborn, M. (2006). Changing the context of teachers' work and professional development: an European perspective. *International Journal of Educational Research*, 45, 242-253.

- Paillé, P. & Muchielli, A. (2013). *L'analyse qualitative en sciences humaines et sociales*. Paris : Armand Colin.
- Panadero, E., & A. Jonsson. (2013). The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review. *Educational Research Review*, 9, 129–144.
- Paquay, L. (2000). Donner du sens à la formation continue. In G. Carlier, J.-P. Renard & L. Paquay (Ed.), *La formation continue des enseignants. Enjeux, innovation et réflexivité* (pp. 263-278). Bruxelles: De Boeck.
- Paquay, L. (2012). Continuité et avancées dans la recherche sur la formation des enseignants. *Les cahiers de recherche du Girsef*, 90, 1-36.
- Paquay, L., Van Nieuwenhoven, C., & Wouters, P. (Dir.) (2010). *L'évaluation, levier du développement professionnel ?* Bruxelles : De Boeck Université.
- Pasquini, R. (2000). Les chefs d'établissement aux prises avec une réforme scolaire : représentations, rôles, compétences et formation. In P.-A. Besencon, R. Pasquini, & C. Petitpierre, *Innovation et vie des établissements scolaires : compétences et rôles des acteurs* (pp. 237-361). Genève : Les Cahiers de la section des sciences de l'éducation.
- Pasquini, R. (2013). *Quand les récits de pratique enseignante parlent d'apprentissage*. Québec : Les Presses de l'Université Laval.
- Pasquini, R. (2016). Et si, pour mieux saisir la complexité de la pratique, nous passions par l'écriture ? *Didactiques en pratique*, 2, 21-29.
- Pasquini, R. (2016, août). *Implementing expanded curricular alignment within teachers' summative assessment practices: issues and prospects*. Communication présentée au colloque EARLI SIG Assessment and Evaluation, Munich. Repéré à: https://www.conftool.com/earli-sig1-2016/index.php?page=browseSessions&form_session=10
- Pasquini, R. (2017a). Evaluation: première question vive de la pratique. *Educateur*, 4, 38-39.
- Pasquini, R. (2017b). Evaluation: deuxième question vive de la pratique. *Educateur*, 5, 36-37.
- Pasquini, R. (2017c). Evaluation: troisième question vive de la pratique. *Educateur*, 6, 34-35.

Pasquini, R. (2017d). Evaluation: quatrième question vive de la pratique. *Educateur*, 7, 34-35.

Pasquini, R. (2017e). En guise de conclusion provisoire : quelques constats pour une évaluation sommative cohérente. *Educateur*, 8, 38-39.

Paterno, J. (2001). Measuring success: A glossary of assessment terms. *Building cathedrals: Compassion for the 21st. century*. Repéré à :
<http://www.angelfire.com/lwa2/bulidingcathedrals/measuringsuccess.html>.

Pelage, A. (2003). La redéfinition du métier de chef d'établissement secondaire : changement statutaire, construction de l'engagement professionnel et épreuves pratiques. *Revue française de pédagogie*, 145, 21-36.

Pellegrino, J. W. (2006). *Rethinking and Redesigning Curriculum, Instruction and Assessment : What Contemporary Research and Theory Suggests*. Washington DC : NCEE.

Pepinster, Ch. (2015). Désintoxiquer les professeurs et toute l'institution scolaire. In M. Neumeyer & E. Vellas (Eds.), *Evaluer sans noter, éduquer sans exclure* (pp. 60-70). Lyon : Chronique Sociale.

Perrenoud, Ph. (1988). Sociologie du travail scolaire et observation participante : la recherche fondamentale dans une recherche-action. In M.-A. Hugon & C. Seibel (dir.), *Recherches impliquées, recherches-action : le cas de l'éducation* (pp.98-104). Bruxelles : De Boeck.

Perrenoud, Ph. (1994). Curriculum : le formel, le réel, le caché. In J. Houssaye (Ed.), *La pédagogie, une encyclopédie pour aujourd'hui* (pp. 61-76). Paris : ESF.

Perrenoud, Ph. (1996). Le métier d'enseignant entre prolétarianisation et professionnalisation : deux modèles du changement. *Perspectives*, XXVI (3), 543- 562.

Perrenoud, Ph. (2002, juillet). *Adosser la pratique réflexive aux sciences sociales, condition de la professionnalisation*. Conférence d'ouverture de l'École d'été des IUFM du pôle Grand Est, Arras. Repéré à :
https://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_2002/2002_22.html

Perrenoud, Ph. (2010). Les processus de (dé)professionnalisation entre savoir, rapport au savoir et contrôle. *Recherches en éducation*, (8), 121-126.

- Piéron H. (1963). *Examens et docimologie*. Paris. PUF.
- Popham, W. J. (2001). Teaching to the Test ? *Helping All Students Achieve*, 58 (6), 16-20.
- Popham, W. J. (2004). All about accountability / Why assessment illiteracy is professional suicide. *Educational Leadership*, 62 (1), 82-83.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48, 4–11.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46, 265–273.
- Poupart, J. (1997). L'entretien de type qualitatif : considérations épistémologiques, théoriques et méthodologiques. In J. Poupart, J.-P. Deslauriers, L.-H. Groulx, A. Laperrière, R. Mayer & A. P. Pires (Dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (pp. 173-209). Boucherville, Québec : Gaétan Morin Editeur.
- Raab, R. (2015). Quelles opérations de recherche dans une démarche inductive inspirée de l'analyse par catégories conceptualisantes ? *Recherches qualitatives*, 34 (1), 122-142.
- Randall, J. & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26 (7), 1372-1380.
- Reason, P. & Bradbury, H. (2008). *Handbook of Action Research : Participative inquiry and Practice*. London : Sage.
- Reeves, D. B. (2004). The Case Against the Zero. *Phi Delta Kappan*, 86 (4), 324-325.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal*, 12 (3), 315–325.
- Rey, B., Carette, V., Defrance, A. & Kahn, S. (2012). *Les compétences à l'école, apprentissage et évaluation*. Bruxelles : De Boeck.
- Rey O. (2014). Entre laboratoire et terrain : comment la recherche fait ses preuves en éducation. *Dossier de veille de l'IFÉ*, 89, 1-26. Repéré à <http://ife.ens-lyon.fr/vst/DA->

Rieg, S. A. (2007). Classroom assessment strategies : What do students at-risk and teachers perceive as effective and useful ? *Journal of Instructional Psychology*, 34 (4), 214-225.

Roach, A. T., Niebling, B. C. & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction and assessments : Implications and applications for research and practice. *Psychology in the Schools*, (45), 158-176.

Robbes, B. (2013). Epistémologies de la pédagogie, relations aux savoirs et à la didactique. *Education et socialisation*, 34, 1-11.

Roditi, E. (2015). Recherches sur les pratiques enseignantes et relations chercheurs-praticiens, *Carrefours de l'éducation*, 1, (39), 55-68.

Roegiers, X. (2010). *L'école et l'évaluation : des situations pour évaluer les compétences des élèves*. Bruxelles : De Boeck.

Rondeau, K. & Paillé, P. (2016). L'analyse qualitative pas à pas : gros plan sur le déroulé des opérations analytiques d'une enquête qualitative. *Recherches qualitatives*, 35 (1), 4-28.

Roth, M., de Pietro & Sanchez Abchi V. (2014). Comment évaluer la lecture en référence au Plan d'études romand ? *forumlecture.ch*, 3, 1-25. Repéré à http://forumlecture.ch/sysModules/obxLeseforum/Artikel/523/2014_3_Roth_et_al.pdf

Royce Sadler, D. (2009). Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1-15). London : Springer.

Royce Sadler, D. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education*, 34, 159-179.

Saada-Robert, M. & Leutenegger, F. (2002). Expliquer/Comprendre : enjeux scientifiques pour la recherche en éducation. In M. Saada-Robert & F. Leutenegger (Eds.), *Expliquer et comprendre en sciences de l'éducation* (pp.7-28). Bruxelles : De Boeck.

Savoie-Zajc, L. (2009). Triangulation (technique de validation par). In A. Mucchielli (Dir.),

Dictionnaire des méthodes qualitatives en sciences humaines et sociales. Paris : Armand Colin.

Savoie-Zajc, L. (2010). Les dynamiques d'accompagnement dans la mise en place de communautés d'apprentissage de personnels scolaires. *Education et Formation*, 293, 9-20.

Savoie-Zajc, L. (2016). L'entrevue semi-dirigée. In B. Gauthier & I. Bourgeois (Dir.) : *Recherche sociale : de la problématique à la collecte de données* (pp. 337-364). Québec, Québec : Presses de l'Université du Québec.

Sayac, N. (2018). Etude des pratiques évaluatives en mathématiques de 25 professeurs des écoles français : une approche didactique à partir de l'analyse des tâches données en évaluation. *Mesure et évaluation en éducation*, 40 (2), 1-31.

Scallon, G. (2007). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.

Schmidt, W. H., Mc Knight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S. et Wolfe, R. G. (2001). *Why schools matter : A cross-national comparison of curriculum and learning*. San Francisco : Jossey-Bass.

Schön, D. (1983). *The Reflexive Practitioner : How Professionals Think In Action*. London : Temple Smith.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4-14.

Silverman, D. (2009). *Doing qualitative research*. London : Thousand Oaks.

Simard, C., Dufays, J.-L., Dolz, J. & Garcia-Debanc, C. (2010). L'évaluation du travail des élèves. In C. Simard, J.-L. Dufays, J. Dolz & C. Garcia-Debanc (Ed.), *Didactique du français langue première* (pp. 139-164). Bruxelles : De Boeck.

Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment*. Thousand Oaks, CA: Corwin Press.

Snoeckx, M. (2008). Entre théories et pratiques: réconciliation autour du paradigme du

praticien réflexif. In Ph. Perrenoud (Ed.), *Conflits de savoir en formation des enseignants* (pp. 125-140). Bruxelles: De Boeck.

Spillane, J. P., & Miele, D. B. (2007). Evidence in practice: A framing of the terrain. In P. A. Moss (Ed.), *Evidence and decision making* (pp. 46–73). Malden, MA: National Society for the Study of Education.

Squires, D. A. (2009). *Curriculum alignment. Research-based strategies for increasing student achievement*. Thousand Oaks : Corwin Press.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77 (3), 238-245.

Stiggins, R. J. (2001). *Student-involved classroom assessment*. Upper Saddle River, NJ. Merrill/Prentice Hall.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83 (10), 758-65.

Stiggins, R. J., Frisbie, R. J., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice* 8 (2), 5–14.

Tessaro, W. (2013). Evaluation et orientation à l'école. *Enjeux pédagogiques*, 21, 8-9.

Tessaro, W. (2013). Améliorer la qualité des pratiques évaluatives des enseignants : une articulation entre formation initiale et formation continue. *Bulletin de la Haute école pédagogique de Berne, du Jura et de Neuchâtel*, 21, 8-9.

Tessaro, W., Jeannet, Ch., Favre Marmet, A. & Pamm Wakley, V. (2010, septembre). *Entre pratiques d'évaluation en classe et textes institutionnels : l'analyse d'un mouvement transpositif*. Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF). Université de Genève : Genève. Repéré à : <https://plone.unige.ch/aref2010/Livret%20des%20resumes%20AREF2010.pdf>

Tierney, R. D. (2013). Fairness in Classroom Assessment. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125-144). Thousand Oaks, CA: SAGE Publications.

- Tierney, R. D. (2015). Altered grades : A grey zone in the ethics of classroom assessment. *Assessment matters*, 8, 5-30.
- Timperley, H. (2011). Le développement professionnel des enseignants et ses effets positifs sur les apprentissages des élèves. *Revue française de pédagogie*, 174, 31-40.
- Tobola Couchepin, C., Sanchez Abchi, V. & Dolz, J. (2016). Pratiques d'évaluation de la production écrite à l'école primaire : outils et régulations. In E. Falardeau, J. Dolz, J.-L. Dumortier & P. Lefrançois, *L'évaluation en classe de français, outil didactique et politique* (pp. 63-84). Namur : Presses Universitaires de Namur.
- Trigwell, K. & Prosser, M. (2014). Qualitative variation in constructive alignment in curriculum design. *High Educ*, 67, 141-154.
- Uwamariya, A. & Mukamurera, J. (2005). Le concept de développement professionnel en enseignement : approches théoriques. *Revue des sciences de l'éducation*, 31 (1), 133-155.
- Vial, M. (1999, octobre). *Modèles et logiques de l'évaluation*. Communication présentée au colloque international « Ethique et qualité de l'évaluation », Reims. En ligne http://www.michelvial.com/boite_96_00/1999-Modeles_et_logiques_de_l_evaluation.pdf
- Vial, M. (2007). *L'accompagnement professionnel ? Méthode à l'égard des praticiens exerçant une fonction éducative*. Bruxelles : De Boeck.
- Vial, M. (2008). *Se former pour évaluer*. Bruxelles : De Boeck.
- Vial, M. (2012). *Se repérer dans les modèles de l'évaluation*. Bruxelles : De Boeck.
- Vinatier, I. & Morrisette, J. (2015). Les recherches collaboratives : enjeux et perspectives. *Carrefours de l'éducation*, 1, (39), 137-170.
- Vinatier, I. & Rinaudo, J.-L. (2015). Rencontres entre chercheurs et praticiens : quels enjeux ? *Carrefours de l'éducation*, 1, (39), 9-18.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education* 30, (3), 749–70.

Walvoord, B. E. & Johnson Anderson, V. (2009). *Effective Grading : A Tool for Learning and Assessment*. San Francisco : Jossey Bass.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6)*. Washington, DC: Council of Chief State School Officers.

Wiggins, G. (1988). *Educative assessment : Designing assessments to inform and improve student performance*. San Francisco : Jossey-Bass.

Wigging, G. & Mc Tighe, J. M. (2005). *Understanding by design*. Alexandria : ASCD.

Winger, T. (2009). Grading what matters. *Educational Leadership*, 67 (3), 73-75.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40 (2), 241-256.

Xu, Y., & Brown, G. T. L. (2016). Teacher Assessment Literacy in Practice: A Reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

Zhang, Z. & Burry-Stock, J. A. (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement in Education*, 16 (4), 323-342.



UNIVERSITÉ
DE GENÈVE

FACULTÉ DE PSYCHOLOGIE ET
DES SCIENCES DE L'ÉDUCATION

Doctorat en sciences de l'éducation

Thèse de Raphaël PASQUINI

Intitulée : «Le modèle théorique de l'alignement curriculaire élargi pour étudier les pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire : enjeux conceptuels et pragmatiques»

*

La Faculté de psychologie et des sciences de l'éducation, sur préavis d'une commission formée par les professeurs : Lucie Mottier Lopez, directrice, FPSE, Université de Genève, Valérie Lussi Borer, FPSE, Université de Genève ; Stéphane Clivaz, Haute école pédagogique, Vaud ; Christophe Ronveaux, FPSE, Université de Genève ; Walther Tessaro, FPSE, Université de Genève ; Catherine Van Nieuwenhoven, Université de Louvain la Neuve, Belgique

autorise l'impression de la présente thèse, sans prétendre par là émettre d'opinion sur les propositions qui y sont énoncées.

GENEVE, le 11 juin 2018

Le doyen :

Pascal Zesiger

Thèse No 704

Numéro d'immatriculation : 84.807.114

N.B. La thèse doit porter la déclaration précédente* et remplir les conditions énumérées dans les « recommandations aux étudiants qui présentent une thèse ».

THESE

1. **A.** Faculté : FACULTE DE PSYCHOLOGIE ET DES SCIENCES DE L'EDUCATION
- Section : Sciences de l'éducation
- Membres du jury de thèse :
Lucie Mottier Lopez, directrice, FPSE, Université de Genève
Valérie Lussi Borer, FPSE, Université de Genève
Stéphane Clivaz, Haute école pédagogique, Vaud
Christophe Ronveaux, FPSE, Université de Genève
Walther Tessaro, FPSE, Université de Genève
Catherine Van Nieuwenhoven, Université de Louvain la Neuve, Belgique
- B.** Nom du candidat : Raphaël Pasquini
- Titre de la thèse : « Le modèle théorique de l'alignement curriculaire élargi pour étudier les pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire : enjeux conceptuels et pragmatiques »
- C.** Lieu et date d'obtention de la licence ou du diplôme :

Licence en sciences de l'éducation, Université de Genève, 1999

APPRECIATIONS ET RECOMMANDATIONS DU JURY

- | | | | |
|----|---|---|-------------------------------------|
| 1. | Le Jury estime la thèse | - Suffisante | ☐ |
| | | - Bonne | ☐ |
| | | - Excellente | ☒ |
| 2. | Le Jury recommande une publication | - Intégrale | ☐ |
| | | - Partielle ou condensée | ☒ |
| | et le procédé technique suivant : | - Publication sous contrat d'éditeur | ☒ |
| | | - Publication dans une revue scientifique | ☒ |
| | | - Publication par des systèmes informatisés | ☒ |
| | | - Reproduction par photocopie | ☐ |

DECISION DE LA FACULTE

- | | | | |
|----|------------------------------|--|--------------------------|
| 1. | Publication : | - Version intégrale (après corrections apportées depuis la soutenance, sous la responsabilité de l'auteur) | ☐ |
| | | - Version partielle ou condensée | ☐ |
| 2. | Moyen technique : | - Publication sous contrat d'éditeur | ☐ |
| | | - Publication dans une revue scientifique | ☐ |
| | | - Publication par offset | ☐ |
| | | - Publication par des moyens informatisés | ☐ |
| | | - Publication par photocopie | ☐ |
| 3. | Dépôt légal minimum : | 2 exemplaires de la version définitive et intégrale, soit photocopiés et/ou édités * | ☐ |



Signature du directeur de thèse



Signature du doyen