

The ASI Series Books Published as a Result of
Activities of the Special Programme on
ADVANCED EDUCATIONAL TECHNOLOGY

This book contains the proceedings of a NATO Advanced Research Workshop held within the activities of the NATO Special Programme on Advanced Educational Technology, running from 1988 to 1993 under the auspices of the NATO Science Committee.

The books published so far as a result of the activities of the Special Programme are as follows (further details are given at the end of this volume):

- Vol. F 67: Designing Hypermedia for Learning. 1990.
- Vol. F 76: Multimedia Interface Design in Education. 1992.
- Vol. F 78: Integrating Advanced Technology into Technology Education. 1991.
- Vol. F 80: Intelligent Tutoring Systems for Foreign Language Learning. 1992.
- Vol. F 81: Cognitive Tools for Learning. 1992.
- Vol. F 84: Computer-Based Learning Environments and Problem Solving. 1992.
- Vol. F 85: Adaptive Learning Environments: Foundations and Frontiers. 1992.
- Vol. F 86: Intelligent Learning Environments and Knowledge Acquisition in Physics. 1992.
- Vol. F 87: Cognitive Modelling and Interactive Environments in Language Learning. 1992.
- Vol. F 89: Mathematical Problem Solving and New Information Technologies. 1992.
- Vol. F 90: Collaborative Learning Through Computer Conferencing. 1992.
- Vol. F 91: New Directions for Intelligent Tutoring Systems. 1992.
- Vol. F 92: Hypermedia Courseware: Structures of Communication and Intelligent Help. 1992.
- Vol. F 93: Interactive Multimedia Learning Environments. 1992.
- Vol. F 95: Comprehensive System Design: A New Educational Technology. 1993.
- Vol. F 96: New Directions in Educational Technology. 1992.
- Vol. F 97: Advanced Models of Cognition for Medical Training and Practice. 1992.
- Vol. F104: Instructional Models in Computer-Based Learning Environments. 1992.
- Vol. F105: Designing Environments for Constructive Learning. 1993.
- Vol. F107: Advanced Educational Technology for Mathematics and Science. 1993.
- Vol. F109: Advanced Educational Technology in Technology Education. 1993.
- Vol. F111: Cognitive Models and Intelligent Environments for Learning Programming. 1993.
- Vol. F112: Item Banking: Interactive Testing and Self-Assessment. 1993.
- Vol. F113: Interactive Learning Technology for the Deaf. 1993.
- Vol. F115: Learning Electricity and Electronics with Advanced Educational Technology. 1993.
- Vol. F116: Control Technology in Elementary Education. 1993.
- Vol. F117: Intelligent Learning Environments: The Case of Geometry. 1993.

Item Banking: Interactive Testing and Self-Assessment

Edited by

Dieudonné A. Leclercq

Service de Technologie de l'Education
Université de Liège au Sart Tilman
Boulevard du Rectorat, 5 Bât. B32
Sart Tilman, B-4000 Liège, Belgium

James E. Bruno

University of California
Graduate School of Education
405 Hilgard Avenue
Los Angeles, CA 90024-1321, USA



Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo
Hong Kong Barcelona Budapest
Published in cooperation with NATO Scientific Affairs Division

Hypermedia: Teaching Through Assessment

Dieudonné Leclercq and Jean-Luc Gilles

Université de Liège, Service de Technologie de l'Éducation, Bâtiment 32,
Sart Tilman, 4000 Liège 1, Belgique
Tel. 32-41-56 20 72, Fax 32-41-56 29 44, e-mail: U017801 at BLIULG11

Abstract: This paper illustrates in which directions hypermedia could enhance educational assessment. Barriers between learning, teaching and assessment disappear. The autonomy of the learner is increased and, consequently, his/her metacognitive activity. The item bank view is progressively transformed into a learning environment that has to be explored and that helps the learner to explore his/her own learning and assessing strategies.

Keywords: Hypermedias, assessment, exploration, detection, learning styles, language learning, MCQ, confidence, note taking.

1 Introduction

It is often claimed that school assessment is artificial, i.e., does not create the real world background in which the learners' ability should be tested.

It is true that test items are too often verbal or symbolic where they should be figurative or behavioral (to use Guilford's classification, 1956, in his famous "Structure of intellect").

Accessibility and practical reasons have kept testing away from realistic setting. Nowadays, the hypermedia approach enriches the educational assessment issues in many respects. In this paper, only four of them will be considered.

1.1 Content specificity

Some assessment situations seem to be meaningful only in real settings or, at least, in a multimedia environment.

Several examples can be provided :

- The capacity to detect dangers should be evaluated only in real settings (where dangers have to be seen or heard ...).

- The capacity to understand a spoken foreign language can be assessed only by means of oral presentation.

- etc.

1.2 Learning experience specificity

It is well known that a source of the difficulty in characterising the taxonomic level (in Bloom's terms) of a question is the fact that the tester ignores what the

past experience of the testee has been. In some respect, it is unfair to ask some question of somebody if it is not known whether he/she has really be exposed to the content.

When the learner has explored a content through hypermedia facilities, the selection of items can be fairer: the channel (the media) by which the content is evaluated can be made mimetic with the channel used during presentation. If not, it means that the learner is put into a situation of media transferability.

Conversely, when the testee knows in advance what the exam will be, he/she can prepare it in a more relevant way, media permitting.

1.3 Merging of learning and assessment

Up to now, a series of barriers have existed between learning and assessment whereas it is well known from formative evaluation principles that evaluation is a kernel part of the learning process itself.

Concretely, at any point of the learning process, the learner should be able to assess him/herself to check

- what his/her needs are (PRE);
- whether he/she understands (PER);
- how well he/she has memorised or transferred (POST).

This "assessment on request" contrasts with the next one, which is automatic.

1.4 The assessment of the learning process itself

In most research settings, focus is placed on the results (the outcomes) of learning whereas the actual process of learning is more rarely assessed.

Obviously, hypermedias, provided adequate reports of decisions and actions are kept, can help the teacher and the learner him/herself to understand the (sometimes largely unconscious) learning strategies sometimes referred to as "cognitive styles" (a wording we will question since it implies that permanent characteristics of the learner are the most important factors affecting learning behaviour).

**

In the next pages, we will illustrate those four issues with specific applications and results.

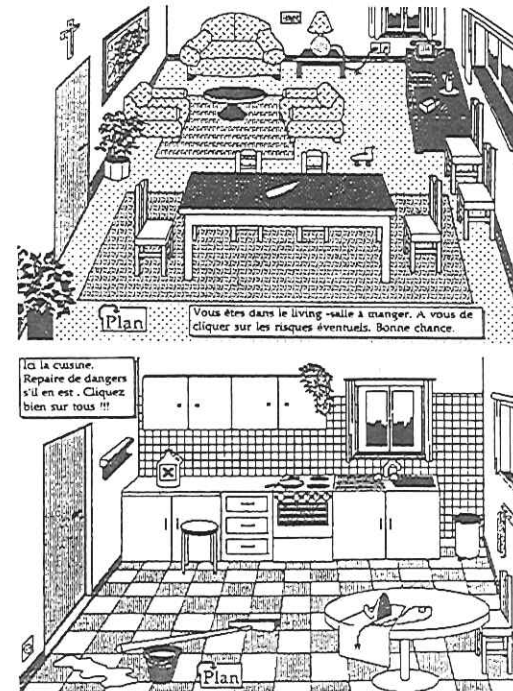
2 Contents Requiring an Hypermedia Approach In Training and Assessment

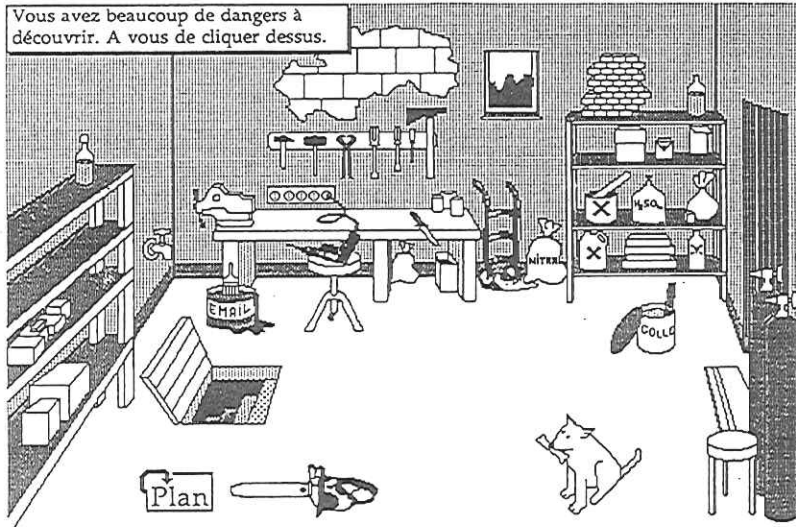
2.1 The DOMUS courseware

Leclercq, Waton and Durllet (1992) have developed hypermedia courseware dedicated to the assessment of the capacity to detect dangers at home. The learner is presented the map of the house and can "walk" from one room to another either by pointing (with the mouse) at a given room on the map, or by pointing an exit door when he/she is already in a given room.

To detect a danger, the examinee has just to point it. The feedback displays reasons why this object or situation is dangerous, so that the feedback is informative and we join here the third issue: the merging of learning and assessment. In cases like this one, it seems relevant to teach through testing, i.e., put the learner in the (complex) situation, have him/her perform and then correct him/her. This should be more efficient for at least two reasons :

1. The learner can express what he/she already knows, avoiding the frustrating position of being told it as if he/she were ignorant.
2. The learner is active from the start, and evolves in successive hypothesis verification moves. Successive performances (after feedback) are likely to "naturally" exhibit progresses, improvements in ability, from simulated "experience".





2.2 The VISPA courseware

Gilles (1991) has developed hypermedia courseware that would be a component in a training program to prevent hazards and accidents in a huge Belgian drinking water production company, namely SPA. The author used Interactive Video to engage the trainees in a task of detecting potential dangers.

The real industrial setting has been video-recorded, from many standpoints (zooms, wide angles, inserts, close ups, travellings, ...) so that the learner can ask different views of the same setting as if he/she walked in it.

This shares common points with the famous Aspen experience where the MIT Architecture Machine Group has put on a videodisk the views of all the streets of a little city (Aspen, Colorado), following each in both directions, with all the possible turns in crossroads and with some "special exploration facilities" such as entering some typical houses (jail, city hall, church, school, etc.) and watching a person's interview (the sheriff, the mayor, the priest, the teacher, etc.).

Consequently, the learner can freely "explore" the town or the industrial setting through the facilities offered by the interface.

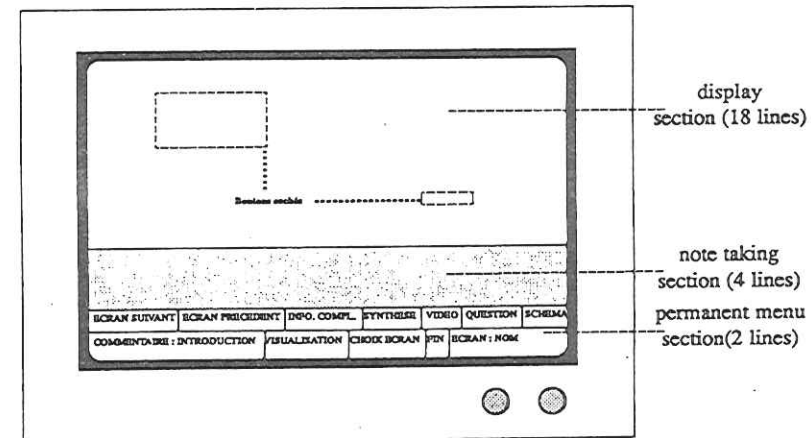
3 Hypermedia to elucidate the learner's learning and assessment experiences

3.1 The DELIN approach

3.1.1 The software

The Service de Technologie de l'Education of the University of Liège has developed a software shell called DELIN (from the French word "delinearisation") to help create hypermedia-type courseware. Figure 1 shows the three important parts of the screen :

- display section (18 lines or icons);
- note taking section (4 lines);
- permanent menu section (2 lines) offering the following possibilities :
 - next screen;
 - previous screen;
 - additional information (actually an additional screen);
 - synthesis of the screen (only keywords organised in a schematic way);
 - video illustration;
 - receive a question to check my comprehension;
 - iconic (still picture) illustration;
 - introduction of notes;
 - consulting previous notes;
 - selection of a specific screen (the codes of which are provided on a paper presenting the "map" of the courseware);
 - end.



3.1.2 The study of variations in strategies

Differential psychology tries to measure interindividual differences (the cognitive style tradition of research). Here, a more "edumetric" approach will consist in studying intraindividual variations of strategies according to variations in constraints and objectives.

Leclercq and Pierret (1989) have introduced two contents (one on Gagne's Taxonomy of Learning, or TAXGAG, and the other one on Chernobyl, TCHERNO, with a computer controlled series of video sequences).

Two groups of 9 students have been asked to learn those two contents, with two different kinds of instructions :

- in instructions a, learners were told that they will have to produce their own summary on what they have learned and present it orally;
- in instructions b, they were told that they will be evaluated through a series of MCQs.

The experimental design was as follows:

	CONTENT	
	TAXGAG	TCHERNO
Group 1 (8)	a	b
Group 2 (8)	b	a
	16	16

3.1.3 Results

Data show that in the instruction b conditions, in the two contents, students adapted

- LESS
 - linear exploration (next screen);
 - requests for additional information;
 - requests for synthesis;
- MORE
 - pointing words to know their meaning;
 - asking for being submitted to MCQs to check their comprehension.

3.1.4 Discussion

By means of a hypermedia tool like DELIN, not only can the learners self assess, but they could see the improvement of their competency after having consulted resources they like. They could even self evaluate whether they improve more when given the information in an iconic format or in a verbal format.

Confidence degrees would be a key feature in this respect since slight changes in the learner's mind would often be noticeable only through changes in confidence level. The current incapacity of measuring those switches from a state of partial knowledge to an other state of (less) partial knowledge keep educationists in the position of biologists who would not benefit from the help of a microscope!

3.2 The AUDIO-SCRIPT experiment

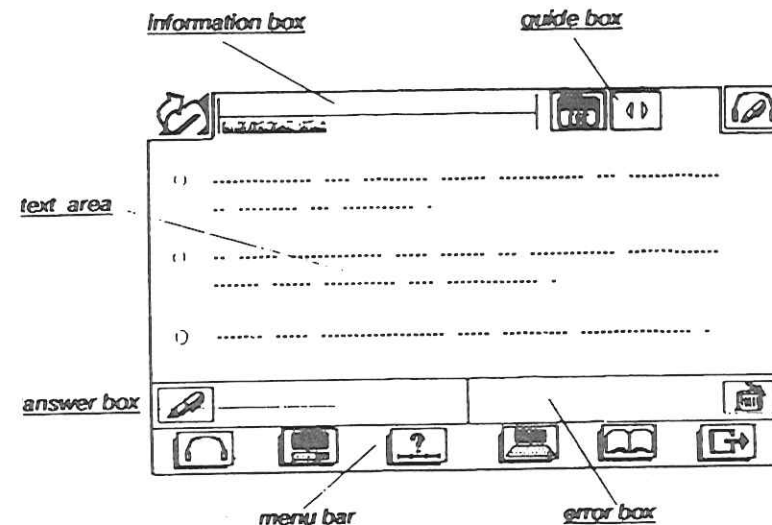
In collaboration with STE-ULg, the SYNAPSE team (Briol, Gilles, Gillet, Kremers and Piette) has developed a software called AUDIO-SCRIPT, that enables a teacher to prepare easily lessons to help learners to understand spoken foreign language from video recordings.

The hardware is a Sony MSX 2 with a genlocker (to superimpose the video image and the computer image on the same screen), a Umatic video recorder. The teacher has to provide a target video sequence (for instance recorded from the BBC TV programs).

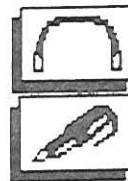
The software helps the teacher

- to delimitate video sequences of a few seconds duration (each sequence containing twenty to forty words);
- for each sequence, introduce (through the keyboard) the oral text so that the computer "knows" its written version;
- introduce a dictionary of his/her own, with definition of each word;
- inform the program of the grammatical nature of words (verb, article, noun, etc.).

This work being done, the learner can start to train in listening oral text. The screen appears with the letters of each word replaced by dots.



Helps offered to students are as follows :

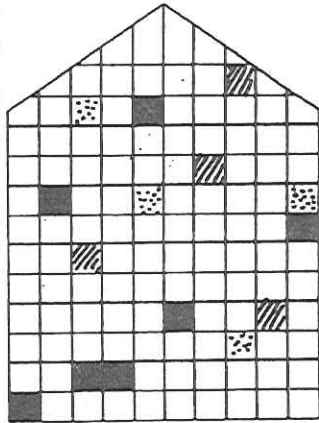


Hear the oral message (without vision).

Type a word he/she has recognized. If it is one of the words really said, the typed word is placed by the program at its correct location on the screen.

Tu as sûrement remarqué que les briques sont de 4 couleurs différentes

- les bleues représentent les questions parfaitement réussies.
 - les vertes représentent les questions bien réussies.
 - les jaunes représentent les questions réussies mais peu sûres.
 - les rouges représentent les questions ratées.
- L'idéal est d'avoir une maison toute bleue.



Appuie sur NEXT

4.2 The free access to EDIP

An interesting issue in free access to item banks, i.e. in using them in a kind of hypernavigational way, is the criterion learners use to select the content on which they will be tested.

Atkinson (1964) has developed an interesting theory about the personal attractiveness of success, i.e. the pleasure different persons experience from success.

According to Atkinson's theory, there are two kinds of persons. The first ones are motivated by searching for successes; the second ones want to avoid failures. This theory predicts that the formers will choose tasks of intermediate difficulty (close to 50 %) whereas the latter will choose either very easy tasks (close to 100 %) that insure to avoid failure, or very difficult ones (close to 0 %) where failure is certain but does not elicit guilt.

Lucassen (1989) has developed an item pool, accessible by computer through a program called EDIP (Boxus and Orban, 1985) on french language at primary level. He observed that his pupils (10 to 12 years old) used to choose easy tasks, as if they wanted to avoid failure, or to experience success. This fits with Atkinson's theory because all those students had previously experienced failures and were likely to try to avoid them.

5 The assessment of the learning process itself

We will develop here only one example: a research on *note taking* behaviour in the DELIN environment.

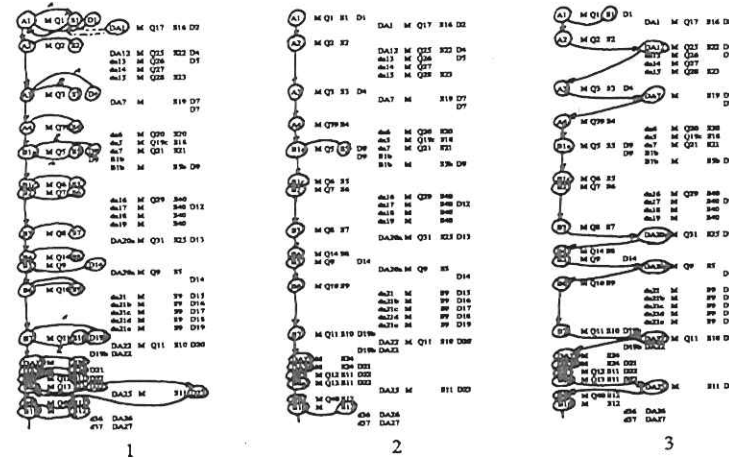
The content

The DELIN "shell" has been filled (Leclercq & Boskin, 1990) with a specific content: Psychological Research on Human Visual Perception.

5.1 The possibility of visualising the strategies

The content is supported by 200 different screens, structured by a "central itinerary" and additional "lateral" branches varying in degree of depth. The figure below shows three different users' personal pathways:

- (1) Continuous requests of synthesis (S) and drawings (D) = visual approach.
- (2) Superficial (global) exploration to get rapidly a general view = holist approach.
- (3) Systematic requests for additional data (DA) = serialist approach.



5.2 The experimental design about note taking behaviour

We wanted to study the behaviour consisting in "taking notes in order to help further consultation of the document", just as one usually does when reading reference books (inserting pieces of paper between 2 pages, folding corners, writing signs in margins, etc.) or when viewing video (noting number count and key words to help remembering sequences).

Graduate school students were asked to explore the content, using freely the possible itineraries and taking notes (on the 4 available lines) with the help of the keyboard. They were told that two weeks after, they would be asked to answer a MCQ test, their electronic notes being made available to them in a paper format.

5.3 The Test: MCQs and confidence

The test contains 15 MCQs, 8 of which been usual (the correct answer is one of the printed alternatives), the 7 others been "general implicit solution" (see Boxus, 1988), i.e. either code 6 (None is correct), 7 (They are all correct), 8 (Lack of data to decide) or 9 (An absurdity in the stem makes the whole question meaningless).

In addition, the students had to provide a confidence degree for each of their questions, on a 6 levels scale (See Leclercq, 1982 and 1988). Tariffs are computed according to decision theory so that students are interested in telling the truth (express their subjectively estimated confidence without bias). Tariffs range from -20 to +20 for each answer (one per question). Mean students scores will be expressed on a 20 points scale, averaged over the 15 questions.

A maximum of 15 screens could be consulted in the testing phase!

5.4 The results (50 students)

a) Annotated screens

Number of screens noted by students during the learning phase vary from 6 to 48, with a mean of 25 (26 by girls, 23 by boys).

Annotated screens are "central screens" (72 %), deepening (15 %), synthesis (10 %) schemata (1,5 %) and questions (1,5 %).

b) Students' opinion

The majority (47 out of 50) of the students commented that the conditions (only four lines, keyboard, severe restriction on characters, ...) made this note taking situation unusual.

c) Recognizing one's notes

Just before the test, the students were invited to retrieve *their own* notes among a series of 10 other ones (all computer printed); 47 students out of 50 were able to recognize their personal notes from the other ones, on the basis of the (declared) following cues : "my abbreviations" (15), "my writing style" (5), "the content" (5), "structure and highlighting" (5), "itinerary" (4), "my genuine spelling errors" (sic) (4).

d) Overall test achievement

Pre-test and Post-test questions were strictly identical. Mean scores (computed with confidence marking tariffs) are : 8,03 on Pre-test; 11,82 on Post-test; that is, a gain of + 3,81.

e) Facilities of groups of questions

For each kind of questions, the objective facility (percentage of correct answers) and subjective facility (average confidence degree) indices were as follows (NQ = number of questions) :

	NQ	Objective Facility			Subjective Facility		
		Pre	Post	Gain	Pre	Post	Gain
Usual	(8)	34	62	+28	47	69	+22
New Reject	(2)	22	58	+36	45	73	+28
All	(2)	52	66	+14	52	78	+26
Lack	(2)	23	30	+7	55	70	+15
Absurdity	(1)	52	64	+12	51	69	+18
Average	(15)	34	58	+24	49	71	+22

f) Students performances according to their characteristics

Students familiar (28) with MCQs gained more than non familiar students (22) : 4,73 instead of 2,64 in average (on a maximum of 20 points).

Students with previous knowledge of the content (13) gain less than more ignorant students (37) : 2,59 instead of 4,19 in average (on 20 points).

Students with negative attitude towards computers (11) gained less than students with neutral attitudes (39) : 2,41 instead of 5,27 in average. Students who had read a book on the content before (4) gained more than students who did not (36) : 4,72 instead of 3,78 in average.

These differences show how sensitive to conditions are the results of a testing.

g) Screen annotation and gain

Consulting annotated screens improves mean number of correct answers for 43 students out of 50 (86 %). Screens have been consulted in 35 % of cases for a correct answer (on Pre-test) and 65 % for an incorrect answer).

h) Screen consultation and lack of confidence

The relation between the confidence degree (on Pre-test) and the frequency of consulting screen at post-tests is as follows :

When confidence degree was	Rate of consulting has been
0 (low)	61 %
1	52 %
2	63 %
3	58 %
4	48 %
5 (high)	26 %

These results support Descartes' view (1636) that "doubt is the incentive of knowledge" (pp. 126-127 in the 1952 edition): the consulting behavior is explained by subjective reasons, not by the "objective" state of our knowledge.

6 Conclusions

6.1 New perspectives for research

As has been seen, students' behaviours are influenced by a series of variables. Knowing the effect of each of them helps interpreting the fundamental variables under study. DELIN (and this kind of software) appears to be a very powerful tool to observe learning strategies, be they conscious or not.

Up to now, researchers were afraid to get too subtle and too numerous data from the learner, since they would not be able to process them. Now, with interactive computer facilities, these large amount of data can be processed *on line*. Realism for instance (see Leclercq, 1993) can be computed on a continuous basis and used in the student model of an intelligent tutoring system (ITS).

6.2 Putting the driving wheel in the learner's hands

Nevertheless, an advance in the direction of the learner's autonomy with self assessment should not be counterbalanced by a backstep by having a "big-brother-like ITS" controlling information presented to the learner. It is his /her choice we want to make more conscious, informed, efficient. If a learner wants to be driven in a tutorial way, let him/her decide it him/herself, as well as coming back to the browsing, free navigating mode.

6.3 Towards new kinds of items

Since getting information (on request) and assessment can be performed by the same channel (the hypermedia), questions will look more and more like case studies: Here is its description. What is your diagnostic? Do you want additional data? In another format? A flavor of that kind of approach can be found in the TASTE methodology (Leclercq et al., 1993).

References

- Ackermans, A., Leclercq, D. & Ernould, C. (1985), Un didacticiel d'évaluation des connaissances sur le diabète chez l'enfant et l'adolescent : MASTER DIAB, pp. 165-178, in Leclercq, Deghaye et Malchair (eds.), Handicaps et Technologies, Liège : Collection Education Santé, Ed. Ceres, Université de Liège
- Atkinson, J. (1964), An introduction to motivation, Princeton: Van Holland
- Boxus, E. (1988), Les QCM à solutions générales au service de l'évaluation à livre ouvert, in: Actes du colloque international "Formation, Evaluation, Sélection par Questionnaires Fermés", Marne-La-Vallée: ESIEE, vol. 1, 318-331
- Briol, P. (1985), "Big Brother Is Helping You", ou "Quand la vidéo interactive donne à Sherlock Holmes l'allure d'un professeur d'anglais new-look", mémoire de licence en Formation des Adultes et Education permanente, Université de Liège
- Brown, G. (1977), Listening to spoken English, London: Longman
- Crosley, K., Green, L. (1989), Le "design" des didacticiels, Paris: A.C.L. Editions
- Descartes R. (1636), Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences
- Descartes R. (1628), Règles pour la conduite de l'esprit
- Gilles, J.L. (1991), VISPA (Vidéo interactive au Service de la Prévention des Accidents du Travail à Spa Monopole). Essai de réalisation en milieu industriel d'un prototype de séquence de formation à la détection de situations de travail "accidentogènes" alliant un ordinateur et un magnétoscope, Mémoire de licence en sciences psychopédagogiques, Université de Mons-Hainaut
- Guilford, J.P. (1959), Three faces of intellect, American Psychologist, 14, 469-479
- Guilford, J.P. (1967), The nature of human intelligence, New York: McGraw-Hill
- Kagan, J. (1965), Impulsive and reflecting children : significance of conceptual tempo, in Krummboltz, J. (ed.), Learning and the educational process, Chicago: Rand McNally, 133-161
- Leclercq, D. (1980), Computerised Tailored Testing : structured and calibrated item banks for summative and formative evaluation, European Journal of Education 15 (3), 251-260
- Leclercq, D. (1983), Confidence marking, its use in testing, in Choppin and Postlethwaite (eds.), Evaluation in education, An international review series, vol. 6, 161-287, Oxford: Pergamon
- Leclercq, D. (1990), Intelligent tutorial and self training system, in Proceedings of the International AI Convention, Nagoya, Nov. 1990
- Leclercq, D. (1991), Hypermédiat et tuteurs intelligents : vers un compromis, in G. Baron & B. de la Passardière, Actes du Colloque Hypermédiat et Apprentissages, Chatenay-Malabry, Sept. 1991
- Leclercq, D. (1992), Psychologie éducationnelle, STE, Université de Liège (5th edition)
- Leclercq, D. and Pierret, D. (1989), A computerized open learning environment to study intrapersonal variations in learning styles: DELIN. In Estes, Heene & Leclercq, New pathways to learning through educational technology, Proceedings of the 6th International Conference on Technology and Education, Orlando, Florida, March 1989, vol. 2, 268-272
- Leclercq, D. and Boskin, A. (1990), Note taking behaviour studied with the help of hypermedia, in Estes, Heene and Leclercq (eds.), New pathways to learning through educational technology, Proceedings of the 7th International Conference on Technology and Education, Brussels: March 1990, 2, 16-19
- Leclercq, D. and de Brogniez, Ph. (1990), A fresh look on confidence marking, in Estes, Heene and Leclercq (eds.), Proceedings of the 7th International Conference on Technology and Education, Brussels: March 1990, 1, 646-649
- Leclercq, D., Boxus, E., de Brogniez, Ph., Wuidar, H. & Lambert, F., (1993), The TASTE approach: general implicit solutions in multiple choice questions (MCQs), Open books exams and interactive testing, in Leclercq and Bruno (eds.), Item banking: Self assessment and interactive testing, NATO ASI Series F, vol. 112. Berlin: Springer-Verlag (this volume)
- Pask, G. (1976), Styles and strategies of learning, British Journal of Educational Psychology 46, 128-148
- Piette, S. (1985), WHY not make a better use of USE ? Essai d'amélioration du didacticiel USE (Understanding Spoken English, 1982) pour des apprenants adultes, mémoire de licence en Formation des Adultes et en Education Permanente, Université de Liège

- Psootka, J., Massey, L.D. & Mutter, S.A., (eds.) (1988), Intelligent Tutoring System: lessons learned. Hillsdale, NY : Lawrence Erlbaum Associates
- Riding, R. & Taylor, E. (1976), Imagery performance and prose comprehension in seven-year old children. *Educational Studies* 2, 21-27
- Schulman, L.S. and Keisler, E.R. (1966), Learning by discovery. Chicago: Rand McNally
- Wauthier, S. (1991), Développement d'une séquence de vidéo interactive pour l'entraînement à la compréhension de l'anglais oral à l'aide du logiciel AUDIOSCRIPT, Mémoire de licence en Sciences de l'Education, Université de Liège
- Witkin, H.A. (1950), Individual differences in case of perception of embedded figures, *Journal of personality* 19, 1-15