A Diagnosis Based on a Qualitative Model of Competence

Stéphanie Jean, Elisabeth Delozanne, Pierre Jacoboni, Brigitte Grugeon

A DIAGNOSIS BASED ON A QUALITATIVE MODEL OF COMPETENCE

Stéphanie JEAN[1], Élisabeth DELOZANNE[1], Pierre JACOBONI[1], Brigitte GRUGEON[2]

[1]*LIUM - Université du Maine, 72 085 LE MANS Cedex 9 (France)*
[2]*DIDIREM - Paris VII, 2 place Jussieu, 75 251 PARIS Cedex 05 (France)*

Abstract: The main focus of the PÉPITE project is how to build a system that helps teachers to assess students in elementary algebra. We describe here prototypes yet tested in actual classrooms and we share our experience of designing practically useful classrooms tools in a participative way. The methodology adopted is a combination of work in mathematics education and user-centred design derived from Human-Computer Interaction research. Starting from a multi-dimensional model of competence in algebra that includes quantitative and qualitative descriptions, we first show how it is applied to analyse both paper-and-pencil tasks and computerised tasks, and secondly how it is applied to analyse the students' productions when performing those tasks. Then we present the prototypes developed to automatically build students' profiles. Finally we discuss the validation process of such an assessment system and our research results.

Keywords: Interface design, evaluation of instructional system, assessment of student's competence, elementary algebra.

1. Introduction

The aim of the PÉPITE project is to develop a tool to help teachers in assessing students' competence in elementary algebra. The 15 year-old students enter French general high schools coming from French college or vocational schools. Most of them encounter strong difficulties and the educational system fails to help them in overcoming those difficulties. As we started this research, our aim was to understand the reasons of such dysfunction, to identify the necessary conditions to a positive evolution, to create appropriate learning situations likely to help evolution of students' knowledge. The idea is to seek out, in the student's way of functioning, the *nuggets* of knowledge (in French, "*pépites*") to use as a basis to build some new knowledge. One of the results of this study is tools enabling teachers to interpret students' production in order to find starting points to modify students' knowledge. The PÉPITE project is the first part of a larger project (not presented here) which aims to assist teachers in choosing activities for students or groups of students corresponding to the starting points highlighted by PÉPITE.

As pointed by Conlon and Pain [3], applied AIED needs "a research methodology that gives a central place to collaboration among teachers, researchers and technologists". HCI research proposes such methods (user-centred design, participatory design, usability engineering) [13] [11]. These methods suggest that users (actual students and teachers) must participate in the design process from the very beginning. We present here our experience in such a multidisciplinary approach and discuss about validation process of assessment system.

In such an approach, the focus is on how to collect *relevant* and *reliable* data with a computer to *make sense* of students' behaviour according to teachers' needs. In this paper we assume that teachers' needs are expressed in a model of competence derived from an educational research presented in section 2. This model gives the kind of results the diagnosis system has to produce (the students' profiles). *Relevance* of the collected data refers to the model. *Reliability* refers to the biases introduced by using a computer and thus refers to interface design problems. *To make sense* refers usually in AI community to diagnosis techniques. It refers also to cognitive and epistemological assumptions about nature of competence in the domain.

Presently in our work we dealt more with difficulty to clarify the model of competence and with interface design problems than with diagnosis techniques. So we focus in this text on how to ensure quality of incoming data that is, in our opinion, of most importance in relation with our objectives.

We here firstly present the educational basis of our work. We then introduce the research objectives of PÉPITE project and the general architecture of the system. We describe each prototype we have implemented and its validation. We point out that difficulties in designing and implementing such a software are not only a diagnosis problem as well known in AIED community, but first an interface designing problem. Finally, we discuss the methodology of validation of PÉPITE and our research results.

2. Educational basis

We begin by presenting what teachers want to know about students, we then present our theoretical framework about mathematical learning and our model of competence in algebra on which our work is based. This section ends with the presentation of the paper-and-pencil diagnosis tool we built.

2.1. What do teachers want to know?

Assessment systems are very often short-item tests consisting of questions that can be answered in less than one minute each. Such systems give a description of student's state of knowledge in term of rates of success / failure. A more popular approach in AIED community bases assessment on student modelling [17]. In those systems the representation of student's knowledge consists of a set of rules, each expressing some small aspect of the domain. This set includes rules for most common misconceptions. A student model is a fine-grained report on student's skills. For example in OLAE [15], the student model reports the probability of mastery of around 290 rules.

Teachers and mathematics educational researchers of our project found inadequate the level of rule mastery to make decisions about elementary algebra teaching. It is not the only dimension of algebra competence.

Let's take an example. Figure 1 shows a student's solution for a classical problem. In term of rules we could say that Karine uses famous incorrect rules:

$x + a \rightarrow x\,a$

$a\,x \pm b \rightarrow (a \pm b)\,x$

$a\,x - x \rightarrow a - 1$

Figure 1a	Figure 1b
A prestidigitator is self-confident while carrying out the following trick. He says to a player: ”Think of a number, add 8, multiply by 3, subtract 4, add the number you thought of, divide by 4, add 2 and subtract the number you first thought of: you have found 7. “ Is this affirmation true? Justify your answer.	$x + 8 = 8x$ $8x$ $3 \times (8x) = 24 + 3x = 27x$ $27x - 4 = 23x$ $23x + x = 24x$ $24x \div 4 = 6x$ $6x + 2 = 8x$ $8x - x = 7$ la solution est bien égal à $\boxed{7}$ [1]

Figure 1a: The prestidigitator problem

Figure 1b: Karine’s paper-and-pencil answer to the prestidigitator problem

Teachers in the PÉPITE project observe three points and then give an interpretation:

– Karine reduces algebraic expressions in order to obtain a result without operator symbol at each right member of an equality. This difficulty is reported by Davis [4] as *process product dilemma*. Nonetheless, Karine’s algebraic formulae keep meaning in relation with the problem and let her use incorrect rules but also correct ones: $3(x+8) \rightarrow 3\times 8x \rightarrow 24 + 3x$.

– Karine translates each sentence of statement into one symbolic expression: Teachers interpret this translation as an algebraic strategy yet very close to an arithmetic one.

– It is possible that knowing the result stirs Karine into using incorrect rules to obtain 7.

Karine has constructed malrules coherent with her conception about algebra as a formal tool to compute a result. In order to help her, it is not efficient enough to show her the right rules. Teachers have proposed to her problem situations involving algebra as a proving tool and emphasising the equivalence meaning of equal sign.

To adapt mathematical activity to student’s state of knowledge, teachers need more than a quantitative description of student’s behaviour. Thus we intended to define a qualitative description in order to help teachers to choose adequate students’ activities.

2.2. Assumptions about mathematical learning

Making sense of learner’s behaviour is closely linked to a theoretical framework about mathematical learning. In this section we present assumptions that found our research.

In order to analyse the dysfunction mentioned above, we feel necessary to define a kind of reference for algebraic competence at this level. We made a synthesis of mathematical, epistemological, didactical and cognitive research works in algebra learning.

According to Douady [5] mathematical concepts have two non-independent dimensions: a tool status and an object status. As far as the tool dimension is concerned competence is expressed in terms firstly of ability to build algebraic expressions and relationships in order to translate (for instance a verbal description of a problem) and to interpret them. Secondly it addresses the ability to choose adequate algebraic tools to solve problems. Different kinds of problems are involved with this tool dimension such as translating problem situations into equations. As far as the object dimension is concerned, we take into account the duality of the algebraic expressions when manipulating them formally: both semantics and syntactic objects. Competence is then expressed in terms of status of algebraic objects, manipulative ability and articulation between their semantic and syntactic attributes linked with other semiotic frames (algebraic, numerical, graphical and geometrical frame and natural language). At this level, we need to consider that algebraic thinking requires a rupture with arithmetic thinking and requires abilities to interpret algebraic expressions both at a procedural and a structural level and to develop a necessary flexibility between the two kinds of interpretations [8] [12] [16].

[1] Translation: “the solution is actually 7”.

2.3. *The multidimensional model of competence in algebra*

Based on this theoretical framework, we have been observing students' behaviour in mathematical classrooms activities during a long period (all the school year round) and have linked those observations with analysis of their exercise books of the previous year [6]. This study highlights that students' productions present coherence and regularities that correspond to their personal knowledge. From this study we kept four dimensions to have a qualitative description of students' algebraic behaviour (cf. figure 2). This model is used firstly to analyse tasks on which students are supposed to learn algebra and secondly to analyse students' productions on those tasks.

<table>
<tr>
<td rowspan="3">From arithmetic to algebra</td>
<td rowspan="3">– using equal sign
♦ announces a result
♦ expresses a symmetric and transitive relation
– calculating with arithmetic numbers
♦ correctly
♦ incorrectly
– using letters
♦ correctly (as unknown to write an equation, as variable to express a relationship or to prove a numerical property)
♦ incorrectly (as generalised number to substitute numerical values, as unspecified to manipulate formulae with incorrect rules, as label or shorthand for a concrete object)
♦ never using</td>
<td>Manipulating algebraic formulae</td>
<td>– good technical mastery
– weak technical mastery (e.g.: not recognising of remarkable identities)
– incorrect technique
♦ bad using of brackets (leading to good / bad result)
♦ using identified malrules
♦ sign errors while transforming
– confusing + et ×</td>
</tr>
<tr>
<td>Translating from a frame to another</td>
<td>– correctly
– correctly but unexpected
– incorrectly (e.g.: square of sum→x^2+y^2)
– abbreviating</td>
</tr>
<tr>
<td>Justifying</td>
<td>– using algebra
– using legal rules
– using formal rules
– arguing in natural language
– using numerical example
– no explanation</td>
</tr>
</table>

Figure 2: Qualitative model of student's algebraic behaviour.

2.4. *The paper-and-pencil diagnosis tool*

Combining this multidimensional model with an ad hoc set of paper-and-pencil tasks we designed a tool enabling teachers to interpret the students' productions in order to establish their profile. This set of tasks has been carefully chosen by researchers and teachers to cover each model dimension. Three types of tasks are proposed to students during a test. *Technical exercises* aim to determine the level of mastery of formal manipulations. *Recognition exercises* aim to determine how students identify and interpret algebraic expressions. *Modelling exercises* aim to identify if students use the expected algebraic type of treatment, how they translate problems into algebraic frame and how they use adapted tools to solve problems.

Matching students' answers to the model provides a diagnosis matrix of values (40x60) linking questions and dimensions of analysis. This very fine description of the behaviour is too detailed to be used by teachers. It is necessary to establish a higher level description: students' *cognitive profiles*. These profiles have three levels of description: a *quantitative description* of algebraic skills in terms of success rates for each type of tasks, a description of *flexibility between algebraic frame and other frames* (represented by a diagram) and a *qualitative description* of functioning coherence.

This paper-and-pencil diagnosis tool has been tested several times. It has in particular been tested in June 1996 on 600 students (21 classes and 7 teachers) of a third form class. This experiment has pointed out that it was difficult and boring for teachers to fill the diagnosis matrix for all their students because the encoding of the students' productions is a very difficult diagnosis task that needs an important didactical expertise. Moreover when several teachers encode same students' tests, diagnosis matrixes may be slightly different but cognitive profiles are in the end identical. It seems to indicate the soundness of the diagnosis tool with respects of teachers' expertise and their acceptance of the algebraic competence description. Furthermore, teachers involved in the experiment are excited at our project to computerise the boring part of the diagnosis.

At last, the students' paper answers coming from this experiment have been used as a corpus for the conception of the PÉPITE project described in next section.

3. The PÉPITE project

The PÉPITE project intends to demonstrate that it is possible to collect with a computer data on students' competence from which experts can build students' profiles, that it is possible to automate this diagnosis and that it is possible for teachers to use these automatically built profiles to make decisions in their classrooms.

Thus, PÉPITE software contains three modules: PÉPITEST collects students' answers to problems adapted from the paper-and-pencil tasks; PÉPIDIAG automatically fills the diagnosis matrix from data collected by PÉPITEST; PÉPIPROFIL from this diagnosis matrix computes the students' profiles and presents them to the users (teachers or researchers).

3.1. PÉPITEST

PÉPITEST is the student interface: it provides problems and gathers students' answers.

In PÉPITEST design, we firstly pay very much attention to *usability problems*. Indeed it is crucial in an assessment environment where collected data had to be interpreted as competence indicators and not to be biased by interface manipulation problems. Ease of learning and short learning times are paramount because students take the test only once. Iterative design is strongly recommended for ensuring interface usability throughout the HCI literature [10] [13]. In [7] we have discussed means to evaluate usability: ergonomic criteria, guidelines, expert walkthrough and pilot test with users [2].

Secondly, we had to create PÉPITEST tasks *as close as possible* to paper-and-pencil tasks in order to get answers *equivalent* to paper-and-pencil ones. Equivalent means that an expert or PÉPIDIAG could interpret them to fill the diagnosis matrix. Let us note that the multidimensional model of competence is used both to diagnose students' productions and to evaluate the PÉPITE tasks. Transferring pencil-and-paper exercises and tools to computational environment is not so obvious. It changes the tasks and has consequences on students' productions. Balacheff [1] calls this *computational transposition*. The main problem in PÉPITEST is that writing an algebraic expression with a pen is very different from typing it on keyboard. From students' point of view, without a specific editor, they have to translate a spatial representation of the expression (e.g. a fraction) into a linear one. From assessment point of view, this translation introduces a difficulty that can disturb diagnose acting as a *distorting mirror* (introducing bias) or that can make visible normally invisible indicators, acting as a *cognitive microscope* [14]. We can propose an algebraic expression editor but it is not yet so easy for student to use it. No PÉPITEST version presented here integrates this editor.

Thirdly, we bear in mind the difficulties in interpreting students' open answers. We could have use form-based user interface allowing students to express their approach without using natural language nor typing algebraic formulae. But it is necessary to allow students to express themselves without monitoring answers, in order to capture for instance their kind of justification or their writing of algebraic formulae. So, we have limited open questions but not to much in order to ensure the test completeness [7].

Presently PÉPITEST runs with 22 problems, with 32 closed questions, 26 answers requiring algebraic expressions and 31 answers using both algebraic expressions and natural language.

As a formative evaluation we first set up a pilot test on October 1996 with 25 students in a high school classroom. As far as the usability was concerned, some minor changes in the test rise to evidence: For instance the basic manipulations (such as carriage return, drag and drop etc.) have to be taught to some students. Nonetheless students have found easy using PÉPITEST 1. As designer, we enjoy that, in spite of difficulties in writing algebraic expressions, students have produced such expressions, and moreover, educational researchers succeed in interpreting them.

PÉPITEST 2 was tested on June 1997 with 43 students in two classes in order to validate PÉPITEST as data collector for diagnosis. Educational researchers in our team were enthusiastic: they were suspicious PÉPITEST would reduce the range of students' productions. For each question, we have found every kind of expected answer proposed in our model of competence in algebra. Thus it shows PÉPITEST *completeness in relation with the model of competence*. In regard to algebraic formulae, as we expected, students had difficulties in producing them. But, those difficulties do not prevent them from answering with algebraic formulae. According to teachers, only one student out of 43 seems to modify her answers. Thus it shows that the expression editor is welcome and useful but may be temporarily bypassed. Finally, educational researchers can fill the diagnosis matrix from students' answers to PÉPITEST problems and the teacher of the class could confirm the profiles thus obtained. So it shows the validity of PÉPITEST in relation with the paper-and-pencil diagnosis tool.

3.2. PÉPIDIAG

PÉPIDIAG is the diagnosis module that analyses answers to PÉPITEST and fills the diagnosis matrix. Closed questions are easy to analyse because we manage to design the interface so that each choice matches expected skills in the competence model. Exercises requiring entering answers with algebraic formulae are more difficult to deal with. Besides linking them with the model skills it is necessary to apply transformations to students' productions in order to normalise them (commutativity, associativeness, etc.). In remaining exercises, in addition to the well-known difficulties processing natural language answers, we face with a segmentation problem, when algebraic formulae occur mixed with natural language.

For this module we presently obtain two main results. Firstly, PÉPIDIAG is able to automatically analyse every closed answer and every simple algebraic expression answer. So we analyse 75 percent of students' answers to PÉPITEST problems. Secondly, we ran PÉPIDIAG on every student's production of our corpus: the system fills the diagnosis matrixes. In order to correlate this partial diagnosis with human assessment, we choose 5 students with different levels of competence and we asked an expert to fill manually the diagnosis matrix. PÉPIDIAG and the human assessor were in agreement. That means that we can already partially automate the diagnosis, but analysing the remaining 25 percent answers still has sense to obtain the completeness of the profiles, particularly concerning the use of letters.

3.3. PÉPIPROFIL

PÉPIPROFIL is the teacher interface: it computes students' profiles and presents them to the teacher. As we explained in 2.4, student's profile has three levels of description: success rates, flexibility between frames, functioning coherences. This results from algorithmic processes merging similar answers and applying thresholds.

We yet obtain two results. With a manually filled matrix, PÉPIPROFIL computes same profile than teacher does. And from the partial matrix yet filled by the system, PÉPIPROFIL builds partial profiles that are confirmed by the teachers. This can be explained by two kinds of reasons: the diagram representing flexibility between frames is very informative and, except for use of letters, the diagnosis gives information for each dimension.

4. Discussion and conclusion

Integrating real teachers and educational researchers in the design team make us focus on student mathematical activities with early field studies. Our approach is very close to PCM Methodology for applied AIED proposed by Conlon and Pain [3] where the development cycle is "driven by practice but informed by, and contributing to theory". This design process requires an early examination of criteria and evaluation methods. Human-Computer Interaction literature recommends that evaluation be considered as a *state of mind* which must express itself throughout the design of a system.

Our validation of PÉPITEST consists in verifying that we obtain equivalent answers in paper-and-pencil test and with PÉPITEST and that data obtained from the software allows experts to build profiles equivalent to paper-and-pencil ones. We evaluate PÉPIDIAG and PÉPIPROFIL by comparing how automatic profiles fit with human assessors ones. It's called *horse-race evaluations* in [18] and *criterion-related validity* in [15].

In regard to student modelling, referring to the model proposed by Balacheff [1], PÉPITEST provides a set of data that is a *behavioural model*. PÉPIDIAG interprets these data to fill the diagnosis matrix, which is the procedural part of the *epistemic model*. Indeed PÉPIDIAG interprets students' productions correlating them with an algebraic skill described in the model of competence. The profile computed by PÉPIPROFIL is a *conceptual model*. Figure 5 shows the matching between PÉPITE and Balacheff's model. Our validation of PÉPITEST corresponds to a *behavioural morphism* between the behavioural models from paper-and-pencil and machines. Agreement between human assessors and PÉPITE demonstrates an "epistemological morphism between the epistemic model and the students' conception as elaborated by research, that is a mapping which preserve the epistemological structures" [1].

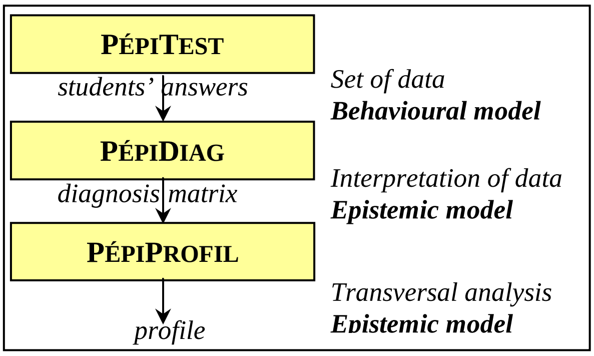

Figure 3:PÉPITE diagnosis tool and Balacheff's model

We intend to develop a tool to help teachers in assessing students' competence in elementary algebra. This tool is not yet completed but we have proved its feasibility. Firstly until the algebraic expressions editor is not integrated, it's not so easy for students to enter such expressions. Secondly, until we progress in interpreting natural language answers, the diagnosis would be partial. Thirdly, we have to make large-scale experiments with teachers to ensure the system acceptability by lay teachers. In spite of these present weaknesses, we have made substantial progress toward our objectives and we have implemented prototypes that already give results that teachers can use. Teachers already recognise their students in the partial profiles made by PÉPITE. Even if important indicators are missing, the partial profiles yet propose a good overall view of students' competence.

The present PÉPITE prototype already gives results that teachers can use. In spite of its still simple diagnosis module, PÉPITE already performs by the quality of incoming data gathered by this interface. This allows us to incrementally develop the diagnosis module, working with corpus obtained from real students' interactions with PÉPITEST. In our opinion this first success is due to teachers' and educational researchers' involvement from the very beginning of the project.

5. References

[1] N. Balacheff, Advanced Educational Technology: Knowledge Revisited, in Liao ed, proceedings of the NATO ASI Serie F. Berlin, Springer Verlag, 1994.

[2] C. Bastien, D. Scapin, Ergonomic Criteria for the Evaluation of HCI, INRIA (RT 156), 1993.

[3] T. Conlon, H. Pain, Persistent Collaboration: A Methodology for Applied AIED, IJAIED (7-3/4), 219-259, 1996.

[4] R. B. Davis, Cognitive Processes Involved in Solving Simple Algebraic Equations, Journal of Children's Mathematical Behavior (1-3), 7-35, 1975.

[5] R. Douady, The Interplay between Different Settings: Tool-Object Dialectic in the Extension of Mathematical Ability: Examples from Elementary School Teaching, in Streefland ed, proceedings of the ninth International Conference for the Psychology of Mathematics Education (2), 33-53, 1985.

[6] B. Grugeon, Design and using of a Multidimensional Structure of Analysis in Elementary Algebra, RDM (17-2), 167-210, 1997. (In French)

[7] S. Jean, E. Delozanne, P. Jacoboni, B. Grugeon, Cognitive Profile in Elementary Algebra: the PÉPITE Test Interface, IFIP-TC3 Official Journal "Education and Information Technology"(3), Chapman & Hall Ltd, 1-15, 1998.

[8] C. Kieran, The Learning and Teaching of School Algebra, Handbook of research on Mathematics Teaching and Learning, Douglas Grouws Ed. Macmillan publishing company, 1992.

[9] K. R. Koedinger, J. R. Anderson, Intelligent Tutoring Goes to School in the Big City, IJAIED (8), 30-43, 1997.

[10] W. Mackay, A.-L. Fayard, HCI Natural Science and Design: a Framework for Triangulation across Disciplines, Designing Interactive Systems, Amsterdam, Holland, 1997.

[11] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, T. Carey, Human-Computer Interaction, Addison-Wesley, 1994.

[12] A. Sfard, On the Dual Nature of Mathematics Conceptions: reflections on processes and objects as different side of the same coin, Educational Studies in Mathematics (22), 1-36, 1991.

[13] B. Shneiderman**,** Designing the user Interface, Addison Wesley, 1992.

[14] M. Twidale, Redressing the balance: the Advantages of Informal Evaluation techniques for Intelligent Learning Environments, IJAIED (4-2/3), 155-178, 1993.

[15] K. VanLehn, J. Martin, Evaluation of an Assessment System based on Bayesian Student Modeling, IJAIED (8-3/4), 179-221, 1997.

[16] G. Vergnaud, Conceptual Fields, Problem-Solving and Intelligent Computer Tools, in De Corte, Linn, Mandl, Verschaffel eds, Computer based Learning Environments and Problem-Solving, NATO ASI serie F.40, 287-308, Berlin, Springer-Verlag, 1992.

[17] E. Wenger, Artificial Intelligence and Tutoring Systems, Computational and Cognitive Approaches to the Communication of Knowledge, Morgan Kaufman Publishers, Los Altos, CA, 1987.

[18] P.H. Winne, A Landscape of Issues in Evaluating Adaptive Learning Systems, IJAIED (4-4), 309-332, 1993.