DEVELOPMENT AND USE OF A DIAGNOSTIC TOOL IN ELEMENTARY ALGEBRA USING AN ONLINE ITEM BANK

Brigitte GRUGEON-ALLYS*, Françoise CHENEVOTOT-QUENTIN**, Julia PILET***, Élisabeth DELOZANNE****

I. Context of the study

This paper addresses the question 2 of the TSG9. Our research concerns the development and use of online resources for diagnostic and differentiation purposes. Teachers’ success in using online banks of exercises, particularly LaboMep from the French organisation Sésamath\(^1\), illustrates that the availability of such online resources for teachers meets a large need (Artigue and al., 2008).

These resources support the teaching and learning of elementary algebra at the end of compulsory education in France (15 years). They take place within the context of multidisciplinary research projects in e-learning environments: the Pépite and Lingot projects (Delozanne et al. 2010). These projects paved the way for the conception and development of the Pépite software, which produces an automatic cognitive assessment for school algebra. This tool, based on a multidimensional analysis of algebraic skills (Grugeon 1997), generates automatic multi-criteria assessments of students’ competence in school algebra (Delozanne and al. 2008) and their cognitive profile. Every profile describes the principal characteristics of a student’s work in algebra.

This paper describes the incorporation of Pépite within the online set of resources LaboMep. Since 2008, our research group, in collaboration with Sésamath\(^2\), has studied the reliability of the diagnostic in providing teachers with appropriate resources for managing a differentiated algebra curriculum that can adequately meet students’ needs. The Pépite experiments suggested that the initial 22-task diagnostic test was too time-consuming. Could a

---

\* Laboratoire de Didactique André Revuz (LDAR) et Université de Picardie Jules Verne – France – brigitte.grugeon@univ-paris-diderot.fr
\** Laboratoire de Didactique André Revuz (LDAR) et Université d’Artois – France – francoise.chenevotot@univ-paris-diderot.fr
\*** Laboratoire de Didactique André Revuz (LDAR) – France – julia.pilet@paris7.jussieu.fr
\**** L’UTES, Université Paris VI, France – elisabeth.delozanne@lip6.fr
\n\(^1\) Since its founding in France in 2001, Sésamath has occupied a central place with teachers among freely available online resources, with nearly 1.5 million visitors in March 2011 on its MathEnPoche (MEP) site. LaboMep is Sésamath’s extension of the MathEnPoche exercise bank.
\(^2\) Projet PépiMEP co-directed by Brigitte Grugeon (LDAR, Université Paris Diderot) and Elisabeth Delozanne (LIP6, UPMC).
shorter test, with 10 well chosen tasks, result in a valid and reliable diagnostic that could be implemented online? How could its reliability be measured? To what extent could such a diagnostic assist mathematics teachers in differentiating students?

We present below the theoretical and methodological elements underlying our approach. We describe in particular the methodology of the individual diagnostic. We reveal the motivations and methodology of the collective class diagnostic, and analyze the data obtained from the diagnostic implemented on the LaboMep platform. We then discuss the principal results.

II. Evaluating 15-year-old students’ algebraic activities: theoretical and methodological elements

Evaluation and diagnostic practices

Diagnostic assessment is a constant part of a teacher’s practice. This term is used to refer to several types of assessment. Ketterlin-Geller and Yavonoff (2009) identify two practices within diagnostic evaluation that use procedures and errors as their basis for analysis: analyses of students' answers to specialized tests and cognitive diagnostic evaluations using standardized and psychometric models. The first, unidimensional, approach lends itself less well to an analysis of the complexity of learning and the relationships between skills and knowledge acquired. The second provides predictive information based on multidimensional attributes that characterize cognitive processes. In both cases, the objective is a local study of students’ misconceptions.

The diagnostic developed through the Lingot project aims to create a global and multidimensional analysis of students’ knowledge and abilities in algebra. It is based on an evaluation of students’ responses (both their errors and procedures) to a series of exercises that is representative of the types of problems encountered in elementary algebra. It does not use psychometric models but is instead based on a cognitive and epistemological study of elementary algebra that enables it to predict students’ learning needs. We describe this study below.

Algebraic activities

Our work is primarily based on two researches. Grugeon (1997) created a model of algebraic competence at the end of compulsory education that could be used as a reference to guide the development of an appropriate diagnostic (Artigue and al. 2002). Using an international synthesis of
research in the didactics of algebra, Kieran (2007) proposed the GTG model of conceptualizing algebraic activities that differentiates three complementary aspects: (1) Generative activities involve producing various algebraic objects: expressions, formulas and equations, and identities, (2) Transformational activities involve the usage of transformational rules (factorization, expansion of products, rules for solving equations and inequalities, etc.), (3) Global/meta-level activities involve the mobilization and usage of the tool of algebra to solve different types of problems (modeling, generalization, proof).

These two approaches allow a categorization of the types of problems encountered in algebra: problems of generalisation and proof, traditional arithmetical problems, problems where algebra appears as a modelling tool, algebraic and functional problems. We use these approaches to define tasks for a diagnostic test and to define different aspects of the multidimensional analysis of students’ activities in elementary algebra.

Beyond research on students’ understanding of the concepts in play, diagnostic evaluations should situate students’ algebraic activity, and their use of techniques in solving problems, along a scale relative to what is appropriate for their grade level.

*Individual diagnostic*

A cognitive diagnostic of students’ abilities in algebra, created from an assessment of their functional knowledge, allows their teacher to make appropriate decisions regarding the management of their teaching and learning.

The diagnostic test included in the Pépite software program is composed of 22 diagnostic tasks (comprising 51 individual items) that cover the range of problems contained within the domain of algebra (Table 1): Exercises involving creating mathematical representations of problems in order to generalize, create a model, complete a proof, or write an appropriate equation (12 items); exercises covering techniques of algebraic calculation (22 items); or exercises in recognition (32 items). The diagnostic tasks may be multiple-choice questions (figure 1) or may be open-ended (figure 2).

In our approach, students’ responses are not only evaluated in terms of correct/incorrect, but also coded in terms of coherences, which are determined *a priori*, that correspond either to appropriate skills and abilities
for the grade level, or to recurring errors. The main characteristics of a student’s cognitive profile, considered relative to their grade level, are calculated through a transversal analysis that codes the solutions of the students to the 22 tasks. The description is both quantitative, in terms of success rate for each type of task, and qualitative, through the dimensions of analysis.

Figure 1 - Justifications suggested if a student selects "false"

Figure 2 - Example of a problem-solving task

Collective diagnostic of a class

The above model provides a description of the cognitive profile of each student. However, teachers would prefer to be able to use the diagnostic to manage classroom instruction. They requested that the diagnostic form groups of students according to their knowledge and skills in algebra, thereby providing a more useful tool for managing the heterogeneous nature of students’ work in algebra.

3The coding of students’ responses to different questions on the diagnostic test is accomplished using a grid of multidimensional analysis (Grugeon, 1997) with five categories: the validity of the response (V), the use of letters as variables (L), the written algebra produced during symbolic transformations (EA), the representations used during translation of a problem (T), and the level of justification (J).
We define a **cognitive stereotype in elementary algebra** (Delozanne et al. 2010) as a set of equivalent profiles that can be considered to be close enough that students can work on tasks with the same learning goals.

We specify the model of a stereotype in elementary algebra along three components: Usage of algebra for solving problems (coded UA); Flexibility in translating different types of representations (geometric figures, graphical representations, natural language) into algebraic expressions, and *vice versa* (coded TA); and ability and adaptability in the various uses of algebraic calculations (coded CA). This model uses similar components to the GTG model of conceptualization of algebraic activity developed by Kieran (2007). For each of the three components, different levels have been identified, along with appropriate benchmarks for each level (Delozanne & al. 2005).

**Using the Pépite diagnostic with Sésamath exercises**

What diagnostic tasks have predictive power? How many diagnostic tasks are necessary to create a reliable and valid diagnostic? The diagnostic prototype used with LaboMeP consists of 10 diagnostic tasks, taken from the original Pépite exercises that are broken down into 27 items. Table 1 presents a comparison of the types of diagnostic tasks contained in the LaboMeP test to those found in the initial test. Our methodology consisted of developing prototypes⁴ that were immediately tested by a small group of expert users, and that were then modified before being made publicly available. The user feedback was then carefully studied.

![Table 1 - Comparison of diagnostic tasks](image)

⁴The definition of generic models (Delozanne, Prévot and al. 2008) permits the creation of “clones” of Pépite exercises, with the same type of problem statement and the same codings.
III. Methodology, description, and analysis of data: validity and reliability of the diagnostic

Methodology, description and analysis of the data

The validity of the diagnostic that was used with the Sésamath item bank was studied using a combinatory analysis (Darwesh 2010) supplemented with a didactic analysis.

(1) A combinatory analysis: It consisted of comparing the obtained stereotypes with both the complete, 22-task test, and with combinations of 15 tasks, using a population of 361 students. Thirteen tasks (1, 2, 3, 4, 9, 10, 11, 12, 13, 14, 15, 16, and 20) were found to appear most often in the best 15-task combinations. Didactic analysis validated the choice of these 13 tasks, and estimated that their number could be further reduced to 10 (tasks 1, 2, 3, 4, 9, 10, 13, 15, 16, and 20). Comparing the stereotypes obtained with the complete 22-task test and the reduced 10-task test yielded an agreement rate of 74%.

(2) A comparative analysis of student groupings in a French seconde class (ages 15-16): The analysis compared four teacher-identified groups to the student groups determined automatically by the LaboMEP diagnostic. Before using this diagnostic, one teacher formed four algebra-learning groups in a seconde class (15-16 years old) of 34 students: students with mastery (group 1), students with approximate mastery but with some difficulties (group 2), average students (group 3), and students with difficulties (group 4). This categorization doesn’t use algebra criteria. The diagnostic completed by Pépite using the online item bank supported the teacher’s choices, while still exhibiting some differences (Table 2). For the 30 students who were present for the diagnostic test, the four algebra learning groups constructed using Pépite were the following:

- Students able to give meaning to algebraic calculations and beginning to develop an intelligent and disciplined use of algebra – CA1 (group A),
- Students whose algebraic calculations were often unmotivated and who often used false rules in their calculations – CA2: using appropriate

The didactic analysis was based on several arguments. The original 22-task test was created with the intention of having redundancies. In this way, the determination of each element of the profile rested on the responses to several tasks, which strengthened the reliability of the diagnostic. However, due to its excessive length, students rarely completed the original test in its entirety, depriving the diagnostic of several rich and informative tasks. Using the a priori analysis of the tasks on the initial 22-task test, we were able to select among tasks of the same type those that had the greatest predictive value.

Research experiments using the online item bank began in early 2011.
algebraic techniques for at least one type of problem (group B); using numerical techniques or inappropriate algebra (group C),

- Students whose understanding of the meaning and usage of algebra is weak – CA3 (group D).

<table>
<thead>
<tr>
<th>Groups formed using Pépité</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Group B</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Group C</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Group D</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>7</td>
<td>12</td>
<td>5</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2- Comparison of groupes formes by teacher and by the Pépite diagnostic in a french seconde class

For twelve out of the 30 students, the groups realized by Pépite and the teacher are the same. An additional 15 of the 30 were placed in higher groups by Pépite than by the teacher.

IV. Discussion of results

Interpretation of results

Our interpretation of these results is based on three factors. First, the experiment took place in the second trimester, after students had already done work in elementary algebra. Second, students were given unlimited time to complete the test and were not judged on their speed in responding to exercises, which was often the case in class. Finally, the interactive computer environment helped students to question their mistakes. The diagnostic also enabled the teacher to identify difficulties in three students out of the 30 whose stereotype-based groups were lower than their initial placement by their teacher. The predictive ability of the LaboMep diagnostic based on the 10-task test is therefore supported by the study of this class. The next step will be to validate it on a larger scale.

Differentiated sets of exercises for each learning group

The automatic LaboMep diagnostic enables groupings within a class. In addition, the diagnostic can assist the teacher in selecting LaboMep exercises appropriate to the needs of the students in each group. This is accomplished using the data on students’ skills in algebraic calculation (CA) and in usage of algebra to solve problems (UA). The following is an example of a possible scenario. A teacher gives her class the LaboMep test before beginning a unit on important mathematical identities. She then decides to
use both the proposed groups and the suggested exercises that correspond to students’ learning needs to structure her instruction. We will consider one group of 15-year-old students in a troisième class (ages 14-15). They cannot give meaning to algebraic calculations without relying on numeric examples, and use misrules of transformation such as concatenation \((4a^3 + 3a^2 \rightarrow 7a^5)\) or false linearity \((a^2 \rightarrow 2a)\). They give little meaning to letters serving as variables and rarely use algebra as a problem-solving tool to produce generalizations, proofs, models, or equations. However, the students can correctly translate mathematical relationships algebraically, as long as no reformulation is necessary. This is a strength that may be exploited. Which objectives should be emphasized in selecting exercises for these students? Here are some possibilities: (i) giving motivation and meaning to the use of letters as representations of letters through exercises that require generalization and proof; for example, by investigating the equivalence of two series of calculations, (ii) destabilizing misconceptions regarding variables, rules of transformation, or translation, through numerical counterexamples or a change of context. For (ii), we propose exercises to destabilize false identities by leading students to describe algebraic and numeric frameworks, using counterexamples to prove that a relation is false (see figure 3).

![Figure 3 - Task for destabilizing false identities](image)

**Conclusion**

We conducted a study of the validity of a diagnostic tool, developed in research laboratories and transferred to LaboMep, an online resource of exercises that is mainly used by mathematics teachers at the collège level (11-15 years). This study was based on both a combinatory analysis and a didactic analysis, both of which are described above. The study shows that the diagnostic has predictive utility in choosing which objectives to emphasize in developing a course of instruction that is appropriate to the needs of different groups of students, as described above in brief. Further large-scale research is still needed.

**REFERENCES**

---

7 Acknowledgements: Thank you to Rebecca Freund for the translation.


