



HAL
open science

Using external assessments for improving assessment practice of primary school teachers: A first study and some methodological questions

Nadine Grapin, Nathalie Sayac

► To cite this version:

Nadine Grapin, Nathalie Sayac. Using external assessments for improving assessment practice of primary school teachers: A first study and some methodological questions. CERME 10, Feb 2017, Dublin, Ireland. hal-01949277

HAL Id: hal-01949277

<https://hal.science/hal-01949277>

Submitted on 9 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using external assessments for improving assessment practice of primary school teachers: A first study and some methodological questions

Nadine Grapin¹ and Nathalie Sayac²

¹ Université Paris Est Créteil, Laboratoire de Didactique André Revuz (Universités Paris Diderot, Paris-Est Créteil, Artois, Cergy Pontoise et Rouen). France; nadine.grapin@u-pec.fr

² Université Paris Est Créteil, Laboratoire de Didactique André Revuz (Universités Paris Diderot, Paris-Est Créteil, Artois, Cergy Pontoise et Rouen). France; nathalie.sayac@u-pec.fr

A recent research (Sayac, 2016) has shown that assessments proposed by primary school teachers are mainly summative and not used to increase student knowledge. To further this work, we have decided to study teachers' assessment practices in mathematics, but also to improve them in the context of a collaborative environment. In such a project, as researchers and teacher educators, we share the same goal with the teachers involved: developing assessment for learning and helping students to learn better in mathematics. For studying this environment and for analysing the professional development of any actor, we use the Activity Theory framework (Engeström, 2001) but also the notions of "evaluative episode" and "professional judgement in assessment", developed by Sayac (in progress) for defining the didactic paradigm of assessment. In our paper, we present our methodology and focus on one aspect of this research: the use of external assessments as tools for improving assessment practices.

Keywords: Assessment practices, external and internal assessment, validity in assessment.

In France, recent modifications in curricula and institutional directions encourage teachers to assess competencies (and not only knowledge), but also to develop, in their classes, "assessment for learning": teachers should consider that assessment is a part of the didactic process and use assessment information for adjusting their teaching strategies. A recent study by Sayac (2016) with some primary school teachers, has showed that assessments in mathematics are mainly summative and that teachers use the results principally for the end of term report. Moreover, assessment tasks are not complex and previously studied before taking the test.

Furthermore, few assessment tools are made available to primary teachers. Until 2012, national diagnostic assessments were organized at the beginning of Grades 3 and 6 and teachers could use them for a diagnostic purpose, but these tests do not exist anymore. In the same time, the number of large scale assessments in primary school has increased in France, but items are not free and only results are published (Brun & Pastor, 2009, Dalibard & Pastor, 2014; Lescure & Pastor, 2008,). So, teachers cannot use items or results directly for their classes and educators can only exploit general results for providing an overview of trends in students' mathematical knowledge or difficulties.

Since 2016, a bank of exercises (with scoring procedure, explanation of the difficulties and propositions of teaching strategies) has been created for helping teachers to assess students at the beginning of Grade 3. It is a commendable initiative, but it raises a lot of questions about these exercises and their use: even if each task is relevant with regard to its assessment aim, teacher must select many of them to elaborate a complete test. How is this selection done and which competencies are finally assessed? Is scoring guide used or not by teachers? How? Etc.

Like other external assessment tools, this could improve teaching and assessment practices, but merely delivering assessment tasks seems not to be sufficient: for designing tests and using results for a better regulation in classes, we think that, in addition to assessment tools, teachers must have knowledge about mathematical notions, teaching them.

Finally, these observations about the bank of exercises have led us to conduct a study to analyze and improve teachers' assessment practice; we suppose that training teachers in assessment will also impact their teaching as a whole. We have chosen to conduct our research in a specific mathematical domain (whole numbers) at the beginning of the elementary school (Grade 1 to 3), in a collaborative environment. After specifying our research and training aims, we explain our theoretical frameworks and methodology for studying this environment and the professional development of any actor and conclude with perspectives.

Research aims and context

This research aims to continuing to explore the assessment practices primary school teachers in mathematics but also to improve them as part of a collaborative research-training. It is undertake in a special network "AeDeP" at FIE (standing for Associated educational Design-experiment Places at French Institute for Education); this type of project is initially based on an educational question shared by different actors (teachers, researchers, school directors, local authorities) and is built for sharing experiences and designing common tools. Our two key issues are:

1. How do primary school teachers assess their students? We focus on test content (which type of tasks do they propose? What techniques are necessary to solve them? Etc.), but also on how teachers design their assessments (what kind of resources they use? What do they do with the results? Etc.)
2. How can such a collaborative research-training improve teachers' assessment practice, and more generally mathematics teaching?

We present in this paper one part of our three-year project, called "EvalNumC2" and we focus on the development of assessment practices in mathematics at primary school. For this part of the project, regular meetings are planned (one per month) with all the actors (ten primary school teachers and us, two educators/researchers) and with different aims depending on the timing of the research. At the beginning, researchers will only collect the tests produced by teachers (without training) and information about teachers' practice; after, external assessments and didactic tools designed by researchers will be introduced.

Theoretical framework

Activity theory for studying professional development in a collaborative environment

We consider that the Activity Theory expanded by Engeström (2001) is a good framework to study this kind of collaborative project of research and training through the objects and the tools used by the subjects, i.e. the teachers involved and us, as mathematics teacher educators/researchers (MTE-Rs). In this framework, the activity of teachers in which we are interested is their assessment practice and the activity of MTE-Rs is to explore and develop these assessment practices within the collaborative environment. We look at the activity of the different subjects in order to produce

results concerning professional development of each one, promoted through this collaborative device (Jaworski, 2006).

Among the tools used in the AeDeP, we have chosen to focus on two specific ones: a list of criteria for studying the validity of test items and tests designed by researchers. In the following we explain why.

Didactic paradigm of assessment

To study the assessment practices of teachers, we adopt the didactic paradigm of assessment developed by Sayac (in progress). In this framework, teachers' assessment practices are studied through the evaluative episodes they propose during the learning process, but also from the evaluative logic of teachers that becomes apparent in the design of the episodes (resources, method, nature of the tests provided), through their professional judgment in assessment and their grading practices.

A number of researchers draw on the notion of professional judgment when considering learning assessment by teachers (Klenowski & Gunn, 2010; Laveault, 2008; Wyatt-Smith, Morgan & Watson, 2002). For them, professional judgment includes both cognitive process and social practice (Mottier Lopez & Allal, 2008), which is not same as a “mechanical gesture of measurement” (Wyatt-Smith et al., 2010), but must be considered as a “flexible dynamic process comprised of middle and final judgments” (Tourmen, 2009). The professional judgment of teachers could be viewed as an act of discernment and as the ability to build intelligibility of the phenomenon of assessment, while taking into account the epistemic, technical, social, ethical and paradigmatic dimensions of classroom assessment practices (Tessaro, 2013). In the didactic paradigm of assessment, the professional judgment is considered as a kind of “didactic vigilance” (Pézard, 2010) specifically applied to the assessment activity of teachers. This allows them to on the one hand give a valid verdict (Chevallard, 1989) about students' mathematical knowledge, individually and collectively, from data collected during the different evaluative episodes. On the other hand this allows them to mutually articulate the different moments of the learning process (especially to connect evaluative episodes to the other moments of the learning process), based on data collected during the different evaluative episodes. This professional judgment in assessment is related to teachers' mathematical and didactical knowledge and assessment skills. It also depends on individual factors as beliefs on learning and assessment as well as professional and personal experiences on assessment (Brady & Bowd, 2006; Di Martino & Zan, 2011; Jong & Hodges, 2015).

Validity of tests

Researchers and teachers, and more generally, all assessment designers, have a same preoccupation about the test: they want to be sure that their tests assess what they should assess and only that. In previous work, we have described a methodology and listed didactic criteria for analysing the validity of an external assessment in mathematics (Gravin, 2015; 2016). We transfer and adapt these principles with two different aims: as researchers, for analyzing the content of internal tests designed by teachers (classroom assessment), and, as educators, for helping teachers to construct their own assessments.

For studying the validity of a test of a mathematical domain, we consider two levels: locally (exercise by exercise) and globally (the test as a whole). From a didactical point of view, the *a priori* analysis of each item is crucial because it gives indicators to guarantee that a task is relevant for achieving its assessment aim. For each item, we realize such an analysis specifying the tool or object aspect, the registers implicated with their possible congruence (Duval, 2006), the types of tasks involved in the resolution, the different techniques (adequate and inadequate through curricula) for solving the problem, the arithmetic problem classes (Vergnaud, 1996), the complexity levels (Sayac & Grapin, 2015). We also take into account the techniques involved in the resolution; if an item can be solved with a technique or a strategy different from the ones expected relatively to his assessment' aims, we consider the item as inconsistent.

For studying and ranking items according to their complexity, we have developed a tool (Sayac & Grapin, 2015) which takes into account three factors. In the first one, the wording and the task context are considered (what is difficult to understand the question?), in the second one, the mathematical knowledge involved in the solving process is studied, and finally in the third one, concerns the level of competency (is the task usual or not, does the student have to take initiative?). For each of these factors, we attribute a degree of complexity between 1 (simple) and 3 (complex). We also observed that discussions between teachers arise during the use of this tool because they do not have the same ideas about the complexity, depending on their teaching or their representations of mathematical notions (Sayac & Grapin 2013). So, this tool seems particularly appropriate to use with the teachers in our project.

On a global level, we study whether the items are representative for the curriculum: have all types of tasks been represented? What are the complexity levels (defined *a priori*) of the items? Are they different or similar? Which registers of representation are involved? When a same type of task is represented by different items, are the effective techniques similar? Etc.

Methodological elements and preliminary results

The Engeström triangle (Engeström, 2001) allows identifying, for each subject, objects that will evolve during the collaborative project of the research and training, through the mediation proposed via the tools, the rules and the division of labor and the communities.

For the teachers involved in the project, the main object is to assess their students. It comes to study specifically, from the collaborative environment, tests designed by researcher as one of the tools used. We will study how these tests could:

- 1- Enhance assessment tasks proposed by the teachers in terms of diversity, complexity and coverage of the mathematical domain.
- 2- Develop teacher's professional judgment through the study of students' answers to these tests and the confrontation between all the teachers during the meetings.
- 3- Work on the coding of students' answers and therefore, on grading.

Studying or using external tests could foster the teachers' professional development, because they will be validly designed from the epistemological and didactical point of view (Grapin & Grugeon,

2015). The evidence of validity will allow us to show how these tests could be relevant for the three points above.

We consider as Johnson, Severance, Penuel and Leary (2016) that:

Professional development organized around the analysis of mathematical tasks has potential to prepare teachers for standards implementation by helping them develop common understandings of standards and how to help students meet ambitious new learning goals. (p. 173)

Therefore, we believe that the contribution of assessment tasks from external tests, in a collaborative context, could develop teachers' skills on assessment tasks design and contribute to enhance their professional judgment in evaluation (Gueudet, Pépin & Trouche, 2013).

Methodology

At the start of the school year, we collected tests designed by teachers involved in our project in order to analyze the assessment tasks with the tool developed in previous research (Sayac & Grapin 2015). We will also collect new tests designed by the teachers at the end of the school year, after the collaborative group work in the meetings. Each assessment will be analyzed in terms of its validity with the criteria listed above; we will principally observe the change in content between the beginning and the end of the project (variety of type of tasks, complexity of tasks and the coding of students' answers).

Each teacher will also fill in a questionnaire about his or her assessment practices (How does he or she design tests? Which resources does he or she use? What are the periods of assessment in his/her classes? How does he or she use the results?). The results of these questionnaires will be used to compare teachers but also as an element for analysing the evolution of their practice.

Lastly, for relying assessment and teaching in classes, each teacher involved in the project will have to keep a "daily book" (journal) in which he or she explains briefly the aim and the content of each mathematical course (in the numerical domain). He or she will also have to identify, according to his/her own representations, the evaluative episodes and describe these more specifically. We will analyze the content of the tasks proposed in tests and during teaching to study their correlations and their evolution during the project. As observed by Grugeon and Bedja (2016), we suppose that training teachers in assessment will also improve teaching: teachers should propose a wider variety of types of tasks, but they also should be able to have a better interpretation of students' errors and propose adapted instruction to upgrade students' level of understanding.

Preliminary results

At the time of writing, we cannot present full results because this project started in September 2016 and we are in the process of collecting first data; we show however two example of tasks, extracted from the same test, one of the teachers in our study used in Grade 3.

In a first task (Figure 1), the five questions are similar and aim to assess the decomposition of written numbers in canonical expressions. In all examples, the underlying structure is regular. We observe that such a task is not complex (we quote level 1 on each factor of complexity) and assesses five times the same knowledge: only the positional aspect of numeration (and not the decimal aspect).

■ **Décompose comme dans l'exemple.** Exemple : $50\ 482 = 50\ 000 + 400 + 80 + 2$

5 472 = $5\ 000 + 400 + 70 + 2$ ✓

8 070 = $8\ 000 + 70$ ✓

56 030 = $50\ 000 + 6\ 000 + 30$ ✓

560 000 = $500\ 000 + 60\ 000$ ✓

205 009 = $200\ 000 + 5\ 000 + 9$ ✓

Figure 1: First exercise and students' answer extracted from a classroom assessment designed by a Grade 3 teacher.

The second exercise of the test (Figure 2) looks like the first one and assesses the same type of knowledge.

Décompose comme dans l'exemple. Exemple : $42\ 500 = (4 \times 10\ 000) + (2 \times 1\ 000) + (5 \times 100)$

4 750 = $(4 \times 1\ 000) + (7 \times 100) + (5 \times 10)$ ✓

35 000 = $(3 \times 10\ 000) + (5 \times 1\ 000)$ ✓

260 050 = $(2 \times 100\ 000) + (6 \times 10\ 000) + (5 \times 10)$ ✓

304 008 = $(3 \times 100\ 000) + (4 \times 1\ 000) + (8 \times 1)$ ✓

540 600 = $(5 \times 100\ 000) + (4 \times 10\ 000) + (6 \times 100)$ ✓

Figure 2: Second exercise and students' answer extracted from a classroom assessment designed by a Grade 3 teachers.

Throughout the full test, there isn't any exercise for assessing the decimal aspect of numeration. We conclude that such a test is not valid and in the project we are going to elaborate other questions with the teachers to fill this lack.

We have not yet achieved the analysis of all classroom assessments designed by teachers involved in the project, but it seems that, as we can observe in the two previous examples, tasks are repetitive, having low complexity and for the numeration, assess principally positional aspect of the numeration. Such observations led us to propose external assessments with other types of tasks: exercises designed to assess numerical aspect of the numeration but also complex situations intended to develop students' abilities.

Conclusion and perspectives

We have focused in the paper on the original and theoretical notion of "didactic paradigm" designed recently by Sayac (in progress). What we can tell currently, from the first data collected (tests, questionnaires, interviews) is that the teachers, participating in our research, design tests with low levels of complexity and have invested very little in assessment as a professional gesture. They assess their students as they can, with very subjective practices. So, it seems that our research, with its training dimension, will make possible a real professional development concerning the assessment tasks proposed in mathematics by these teachers and their professional judgment in assessment.

So, besides studying and providing primary school assessment practice in mathematics and designing assessment tools, our research aims also to develop these theoretical elements. At the end of the project, it will be possible to show the impact on the teachers' practice and interests, and the limitations of this methodology, according to our theoretical framework. It would then be possible to extend such studies in other mathematical domains (as geometry) or in other levels, for example at secondary school.

References

- Brousseau, G. (1997). *Theory of didactical situations in mathematics* (edited and translated by N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield). Dordrecht, NL: Kluwer.
- Charles-Pézard, M. (2010). Installer la paix scolaire, exercer une vigilance didactique. *Recherches en didactique des mathématiques*, 30(2), 197–261.
- Chesne, J-F. (2014). *D'une évaluation à l'autre : des acquis des élèves sur les nombres en sixième à l'élaboration et à l'analyse d'une formation d'enseignants centrée sur le calcul mental*. Université Paris-Diderot, Paris.
- Dalibard, E., Pastor, J.-M. (2015). CEDRE 2014 - Mathématiques en fin d'école primaire : les élèves qui arrivent au collège ont des niveaux très hétérogènes. *Note d'information*, 18, MEN - DEPP.
- Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: a bridge between beliefs and emotions. *ZDM*, 43(4), 471–482.
- Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, 61(1-2), 103–131.
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14, 133–156.
- Gueudet, G., Pepin, B., & Trouche, L. (2013). Collective work with resources: An essential dimension for teacher documentation. *ZDM*, 45(7), 1003–1016.
- Grapin, N. (2015). *Étude de la validité de dispositifs d'évaluation et conception d'un modèle d'analyse multidimensionnelle des connaissances numériques des élèves de fin d'école*. Université Paris Diderot. Paris.
- Grapin, N. (2016). *Validity of mathematics large scale assessment: a didactic analysis*. Paper presented at the 13th International Congress on Mathematical Education. Hamburg, 24–31 July 2016.
- Jaworski, B., & Goodchild, S. (2006). *Inquiry community in an activity theory frame*. In J. Novotna, H. Moraova, M. Kratka & N. Stehlikova (Eds.), *Proceedings of the 30th conference of the international group for the psychology of mathematics education* (vol. 3, pp. 353–360). Prague: Charles University.
- Johnson, R., Severance, S., Penuel, W.R. & Leary, H. (2016). Teachers, tasks, and tensions: lessons from a research–practice partnership, *Journal of Mathematics Teacher Education*, 19 (2-3), 169–185.

- Jong, C., & Hodges, T. E. (2015). Assessing attitudes toward mathematics across teacher education contexts. *Journal of Mathematics Teacher Education*, 18(5), 407–425.
- Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue suisse des sciences de l'éducation*, 30, 483–500.
- Morgan, C., & Watson, A. (2002). The interpretative nature of teachers' assessment of students' mathematics: Issues for equity. *Journal for research in mathematics education* 33(2) 78–110.
- Mottier Lopez, L. & Allal, L. (2008 / 3). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 3, 465–482.
- Sayac, N. (2016). *Nature et diversité des tâches mathématiques proposées en évaluation sommative par des professeurs des écoles, en France*. Paper presented at the XXVIIIème Colloque de l'ADMEE- Lisbonne-13-15 janvier 2016.
- Sayac N., Grapin N. (2013). *Former à l'évaluation à partir d'un outil de chercheur : enjeux et perspectives*. Paper presented at the XXXVIII Colloque Admee – Fribourg.
- Sayac, N., Grapin, N. (2015). Évaluation externe et didactique des mathématiques : un regard croisé. *Recherches en didactique des mathématiques*, 35 (1), 101–126.
- Tessaro, W. (2013). Améliorer la qualité des pratiques évaluatives: une articulation entre formation initiale et formation continue. *Enjeux pédagogiques*, 21, 8-9.
- Tourmen, C. (2009). Evaluators' Decision Making The Relationship Between Theory, Practice, and Experience. *American Journal of Evaluation*, 30(1), 7–30.
- Vergnaud, G. (1996). The theory of conceptual fields. In L. Steffe & P. Nesher (Eds), *Theory of mathematical learning* (pp. 219-239). Mahwah, NJ : Lawrence Erlbaum Associates.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), 59–75.