13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

FORMATIVE ASSESSMENT IN MATHEMATICS EDUCATION BY USING TECHNOLOGY

Marja van den Heuvel-Panhuizen[ab], Ilona Friso-van den Bos[b], & Mieke Abels[a]

[a]Freudenthal Institute, Faculty of Science & [b]Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands

This paper reports about the Dutch part of the FaSMEd project (Formative Assessment in Science and Mathematics Education). FaSMED is a EU-funded project in which nine institutions in eight countries investigate how technology in the classroom can be used for formative assessment. To gain knowledge about this the Dutch team developed a Digital Assessment Environment (DAE) for formative assessment in upper primary school mathematics education. The DAE contains four mathematics domains, including percentages. To test the usefulness of the DAE, an experiment has been carried out in sixth-grade classrooms in 22 schools in the Netherlands. The present paper reports on this large-scale experiment in which it is investigated how the DAE facilitated formative assessment, and whether using the DAE had an effect on the students' mathematics performance.

INTRODUCTION

Formative assessment

Good education requires that the given instruction has to fit the students' level of understanding. Therefore, the teacher should know both which obstacles there are in their students' learning, and what will help their students' learning process. Formative assessment can provide teachers with this knowledge. According to Black and Wiliam (2009, p. 9) an assessment practice can be defined as formative "to the extent that evidence about student achievement is elicited, interpreted, and used […] to make decisions about the next steps in instruction." Therefore, formative assessment is considered 'assessment *for* learning' (e.g., Wiliam, 2011) and contrasted with 'assessment *of* learning' which refers to summative assessment that aims to evaluate a student's learning at the end of an instructional sequence to give him/her a mark or a certificate. Formative assessment is an interim-assessment intended to find clues for further instruction. In fact, it is something that teachers continuously do during their teaching.

Teachers can use a wide range of methods to collect information about their students' learning (e.g. Veldhuis, Van den Heuvel-Panhuizen, Vermeulen, & Eggen., 2013); for example, asking their students questions, observing them when they are working alone or in small groups, or having them do a series of teacher-made assessment tasks. However, externally developed standardised tests and textbook tests can also be employed for formative assessment, at least as long as the goal of using them is making didactical decisions about how to proceed with further instruction. In other words, it is not the format that determines the difference between formative and summative assessment, but the intention with which an assessment method is used. For a test to be used formatively, it needs to provide more than just a total score of correctly answered problems.

Because studies have shown that formative assessment can raise levels of student achievement (Black & Wiliam, 1998) and because teachers' assessment skills are positively associated with students' mathematics achievement scores (e.g., Christoforidou et al., 2014), more and more support has been given to the implementation of formative assessment in the educational practice over the two last decades (OECD/CERI, 2008).

The role of technology in formative assessment

When digital techniques entered schools, this led to the question whether the use of computers and other electronic devices, and the access to internet might contribute to teachers' formative assessment practices. For example, according to Quellmalz (2013) technology can support formative assessment in several ways, ranging from making resources and pools of assessment tasks available for teachers, to enabling the assessment of more complex competences. Furthermore, ICT-based assessment can make an assessment interactive and dynamic, can provide individualised feedback and help to students, and can track students' learning (see also Looney, 2010). Similarly, Van den Heuvel-Panhuizen et al. (2011) showed that that ICT-based assessment can enable: (a) the use of high-demand tasks, (b) accessibility of tasks to students, and (c) visibility of students' thinking and solution processes. Especially the latter makes the use of ICT in assessment a very promising direction for improving the formative assessment practice of teachers.

The FaSMEd project

FaSMEd is an acronym for *F*ormative *A*ssessment in *S*cience and *M*athematics *Ed*ucation. The FaSMEd project is a large, international development and research project in which universities from England, Ireland, Germany, Norway, France, Italy, South Africa, and the Netherlands participate. The FaSMEd Consortium is led by the University of Newcastle Upon Tyne, UK. The project is financed by the European Union and aims at researching the use of technology in formative assessment classroom practices in ways that allow teachers to respond to the emerging needs of (low achieving) students in mathematics and science so that they are better motivated in their learning of these important school subjects.

The Dutch FaSMEd project is conducted by Utrecht University. In this Dutch part of the project, the Digital Assessment Environment (DAE) is developed for mathematics education in grades 5 and 6 of primary school (10-12 year-old students). The main characteristic of this DAE is that it is not limited to delivering test scores, but also offers teachers a window into their students' strategies to solve mathematical problems. As such, the DAE provides the teacher with information that subsequently can be taken as a stepping stone for further instruction.

The current study

To gain knowledge about whether a technology-enhanced formative assessment such as the DAE is useful for teachers and can contribute to raising students' achievement level in mathematics, a large-scale experiment was set up that ran from September 2015 to February 2016. The research questions to be answered through this experiment were: (1) How does the DAE facilitate the formative assessment carried out by the teachers? (2) Does using the DAE has an effect on the students' mathematics performance? (3) Is there a relation between the teachers' use of the DAE and the students' change in performance?

METHOD

Participants

Participants were teachers and their sixth-grade students from 26 classes within 22 primary schools. The participating schools were located all over the Netherlands. At the start of the study, informed consent was obtained from the parents or caretakers of the students.

The Digital Assessment Environment

The DAE is an online environment with which teachers can collect information about mathematical skills of their students. The DAE has been built within the Digital Mathematics Environment (Boon, 2009). The DAE contains assessment modules on four mathematics domains: percentages, fractions, metric system, and graphs. For each mathematics domain there are two parallel tests, each consisting of a series of six or seven problems. The perspective that is taken in the DAE is that formative assessment that informs teachers about what students can do rather than what they cannot do, helps teachers to find didactical ways to extend students' understanding and skills. In particular, knowledge about how students proceeded to reach their answers is considered an important point of action for further instruction. Therefore, in the DAE each problem is enriched with a number of digital auxiliary tools (see Figure 1) such as scrap paper, a number line, a bar, a ratio table, or a hint which students can use to complete the problems. When a student clicks on the icon of an auxiliary tool a window with that tool, for example, the interactive bar (see Figure 2), pops up.

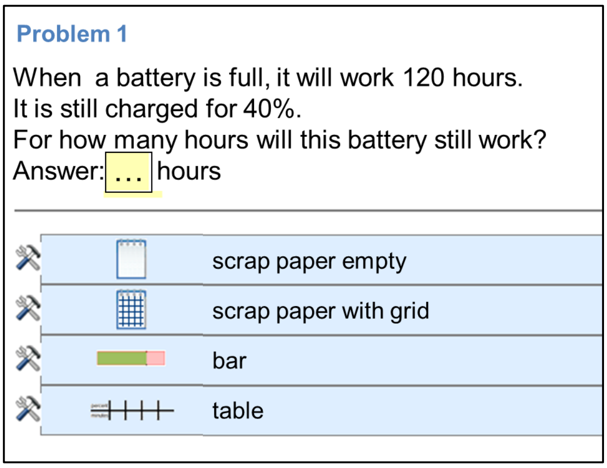

Figure 1: Percentage problem with auxiliary tools in the DAE.

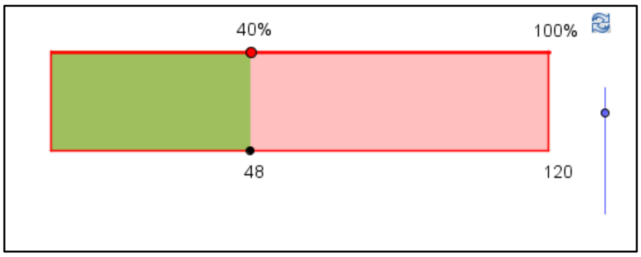

Figure 2: Interactive bar as auxiliary tool.

What auxiliary tool students choose and whether they anyway make use of it is optional. The students are not required to do this. However, by offering students the opportunity to use such a tool, they get the chance to show what they are capable of with some support. In this way the DAE reveals the students' 'zone of proximal development'. Seeing how students employed the auxiliary tools is useful for the teacher to take further steps in instruction.

Professional development

All teachers were invited for three professional development meetings. In the first meeting, the purpose of the study and the idea of formative assessment was addressed. Furthermore, the modules on percentages and fractions were introduced. Between the first and the second professional development meeting, the teachers worked with these modules, and in the second meeting, their findings and their use of these findings were discussed. In addition, in this second meeting the modules on metric system and graphs were introduced. In the third meeting the teachers' findings concerning these modules and use of these findings were again discussed. The third meeting was concluded with a general reflection on the usefulness of the DAE.

Measures

To record how the DAE facilitated the formative assessment carried out by the teachers, they were asked to fill in a questionnaire each time their students completed a test. This means that for the four modules in total, every teacher had to fill in at least four questionnaires. The questions concerned what the teachers had learned about their students, whether they were planning to use this information in further instruction, and if so, how.

To measure the students' mathematics performance, the scores of the Cito Mathematics End Grade 5 test and the Cito Mathematics Mid-Grade 6 test were collected. These served as pretest and posttest scores respectively.

Data analysis

For answering the question how the DAE facilitated the formative assessment, our first focus was on getting to know what the teachers learned from their students' work in the DAE. Hereafter, we looked at whether the teachers used or were planning to use this information for making decisions for further instruction. Teacher responses were also evaluated for quality in terms of specificity and rationale of their formative assessment activities (Polly et al., 2015). To investigate whether using DAE has an effect on the students' mathematics performance, gains in the Cito Mathematics scores of the participating students were compared to the national average gains in these scores.

In the current paper we only report about the data available by mid-February 2016. This means that only the data from the questionnaires about the tests on percentages and fractions were considered. Moreover, only the responses of those teachers were considered who had returned the questionnaires on both the percentages test and the fractions test. This amounted to 40 filled-in teacher questionnaires filled in by 20 teachers from 18 schools. Only from 10 schools the Cito data were available including the data of 218 students in total.

RESULTS

How did the DAE facilitate the formative assessment?

Table 1 shows of all teachers together what the questionnaires revealed with respect to what the teachers learned about their students' understanding through the use of the DAE. We decided to classify the teachers' responses in five categories (A-E) ranging from nothing was learned to a lot was learned. In the end we did not find responses in which teachers indicated that nothing was learned. Furthermore, we discovered only a small number of responses (10%) that made clear that teachers only saw things that were already known. All the other responses indicated that new information was gained from the DAE. However, it should be noted that these findings were based on a liberal coding procedure. The reason for this was that in many responses teachers reported not have learned much from the DAE, but at the same time their responses gave very in-depth information about what they found in the students' work. So, for such responses we decided to allocate them into one of the last three categories (C-E) included in Table 1.

	Frequency of response	
What was learned from the DAE	n	%
A. Nothing was learned	0	0%
B. Only existing knowledge about student performance was learned	4	10%
C. One of three things was learned: what students could or could not do, how students approached something, and which problems were difficult.	10	25%
D. Two of three things were learned: what students could or could not do, how students approached something, and which problems were difficult.	18	45%
E. All of three things were learned: what students could or could not do, how students approached something, and which problems were difficult.	8	20%
Total	40	100%

Table 1: Absolute and relative frequencies of responses which indicated that particular things were learned from the DAE.

An example of what a teacher said she had learned from the students' work in the DAE is given in the following quotation:

> Because of this test I got a better picture of the skills of the students regarding percentages, and I was surprised about the scores of some students. I also saw that the auxiliary tools are used frequently, which puts me in a positive mood. Of the students who obtained a score of 0%, none used the auxiliary tools. This is something I can work with. *(translation into English by authors)*

This teacher discovered the potential of auxiliary tools, and how these tools could support the process with which her students tackled problems with regard to the domain of percentages. She was also surprised by the scores of some individual students, indicating that she had learned something about these students specifically. Therefore, the response was coded within the category D.

Of the 20 teachers whose questionnaires we analysed, 16 indicated that they had learned a lot (C-E) from both the percentages and the fractions test. Only four teachers indicated that they had learned little (A-B) from one of the tests: two indicated that they learned little after analyzing the test on percentages, and two indicated that they learned little after analyzing the test on fractions. No teachers indicated that they had learned little on both of the two tests.

Regarding the use of the information obtained through the DAE for instructional decision making, of the 20 teachers there were three teachers who stated on at least one of their two questionnaires that they had no intention to use the information. One of these teachers reported that he did not intend to use the information on either the percentages or the fractions test, while the other two only had no plans to use the information obtained with the percentage test, but did intend to use the information obtained with the fractions test. In total, six teachers had plans to use the information on one of both tests. Furthermore, 11 teachers reported that they had the intention to use the information obtained using both tests, but did not yet carry out their plans. Finally, four teachers reported that they had already started using the information obtained through one of the tests, and two teachers had already started using the information obtained using both tests, for example by providing additional instruction on difficult topics or locating difficulties in solving the test problems with students. Remarkably enough, three teachers who reported that they learned new things from their students' work in the DAE (for example, that the students did not read the problem well), also reported that they did not intend to use the information obtained.

The way in which the teachers reported about their findings and how they would use them differed a lot. Some teachers specified no plan, or only stated without further specification that the information would be used. This was true for four teachers' responses to both questionnaires and for one teacher's response to the questionnaire about the fractions test only. A more elaborate reaction, but still in global terms (e.g. *'I want to address this domain more')*, was given by one teacher to both questionnaires, and by two teachers to only the questionnaire about the percentage test. A third category of responses were those in which it was specified either which students would be targeted, or the approach that would be taken. This was the case for seven teachers' response to one of the two questionnaires, and for six teachers' responses to both questionnaires. The last category included responses with the highest level of specificity in which both the targeted students and the selected approach were specified. Of eight teachers, their responses to one questionnaire could be allocated to this category. An example of such a response is:

> As duo-colleagues we need to discuss this. As mentioned before, this is a weak group in terms of mathematics. Concentration and sloppiness play a large part in this. 1. The auxiliary tools are used too little. First, I would like to oblige a number of students to always use scrap paper, even if it is only to offload memory. I tell students this on an individual basis, but apparently this is not enough. 2. With 3 to 4 students we really need to get back to the foundations of fractions [*Names removed by authors*]. The fourth student is very poor at mathematics in any case. With percentages, she did not give a single correct answer either [*Name removed by authors*].

Also with respect to the quality of their rationale for making use of the information the teachers differed considerably. Five teachers, gave in both questionnaires no rationale for any plans that were made with the information obtained through the DAE, and another five teachers gave no rationale for their plans in one of their questionnaires. In all other cases, teachers did provide one or more rationales for their plans. Two teachers based their rationale on the total number of correct answers given by children in both questionnaires, and five teachers based their rationale on the number of correct answers in one of the questionnaires. Five teachers gave in both questionnaires a rationale for their plan based on a specific result, such as a specific problem or a specific mistake made by multiple children. Eight teachers gave only in one questionnaire a rationale which was based on a specific result. Finally, one teacher gave in two questionnaires a rationale based on specific didactical knowledge, including knowledge of the end-of-year requirements with regard to a certain type of problem. Four teachers gave a rationale based on a specific result only in one questionnaire.

Did using the DAE have an effect on the students' mathematics performance?

For the 10 schools (218 students) of which mathematics test scores could be collected, we found that the average ability score, 107.97 ($SD = 11.55$), on the Cito Mathematics End Grade 5 test (collected prior to using the DAE) differed significantly from the average ability score, 114.11 ($SD = 10.80$), on the Cito Mid-Grade 6 test (collected after the DAE was used); $t(228) = -6.15$, $p < .001$. However, this growth in achievement score is comparable with the difference between these two measurement points in a national reference sample (average score End Grade 5 is 105 and the average score Mid-Grade 6 is 112) (Flik, 2016); the students in our study did not show larger achievement gains, $t(228) = -1.82$, $p = .07$.

CONCLUSION AND DISCUSSION

The current study was designed to gain knowledge about whether a technology-enhanced formative assessment tool such as the DAE is useful for providing teachers with knowledge about their students' understanding and clues about how to proceed with their lessons, and whether the use of such an assessment tool can contribute to raising students' achievement level in mathematics. Results of the study show that almost all teachers learned a lot from using the DAE. Only four out of twenty teachers learned little from it. Also regarding the use of the information gained by the teachers, the responses evidently showed the DAE's potential. In 70% of the responses it was made clear that the teachers had planned to use the information for instructional decision making and in 20% of the responses it was indicated that this had already started. However, there was large variety in the way the teachers reported about their findings and described their plans for further instruction. Some responses only contained a very global description of the teachers' findings whereas other responses showed intricate and insightful analyses of student work. With respect to the use of the information for instructional decision making there were also large differences in the responses. Some of them revealed that there were no plans to use the information and other responses brought to light that elaborate and specific plans were made to adjust instruction to the needs of classes and individual students.

Students of teachers participating in the experiment did not make greater gains in mathematics than their peers in other schools. This may indicate that the use of the DAE did not prompt teachers, on average, to improve their teaching sufficiently to significantly ameliorate student outcomes. In any case, this short period of working with the DAE to fine-tune teaching to students' needs and find stepping stones for fostering students' learning did not turn out to be more or less effective in raising students' score on the Cito Mathematics test than not using the DAE.

The rich data rendered by the teacher questionnaires, describing what teachers had learned through the DAE from their students' understanding and how they intended to use this information, suggested that technology-enhanced formative assessment is a promising avenue in the field of educational research, practice, and policy. Although the current study was so far unable to show the effects of the use of a formative assessment tool on students' achievement, it did show how much teachers could benefit from using technology-enhanced formative assessment. More extensive analyses will be done after data-collection has been completed, which may provide a better overview of the outcomes of the study.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5,* 7-74. DOI: 10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability, 21*(1), pp. 5-31.

Boon, P. (2009). A designer speaks: Peter Boon. *Educational Designer, 1*(2). http://www.educationaldesigner.org/ed/volume1/issue2/article7/

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation, 40,* 1-11. DOI: 10.1016/j.stueduc.2013.11.006

Flik, A. (2016). *Tabellen tussenopbrengsten Cito LOVS* [Tables intermediate results Cito LOVS]. Retrieved from http://www.hco.nl/images/paginas/werkenaanopbrengsten/OGW_en_data-analyse/2016_Tabellen_tussenopbrengsten_groepsniveau_CITO_LOVS.pdf

Looney, J. (2010). Making it Happen: Formative Assessment and Educational Technologies. Thinking Deeper Research Paper No. 1, Part 3. Promethean Education Strategy Group. http://www.prometheanworld.com/rx_content/files/PDF/MakingitHappenFormativeAssessmentEducationalTechnologies-169721.pdf

OECD/CERI (2008). Assessment for learning. Formative assessment. OECD/CERI. http://www.oecd.org/site/educeri21st/40600533.pdf

Quellmalz, E. (2013). Technology to support next-generation classroom formative assessment for learning. San Francisco: WestEd. http://www.wested.org/wp-content/files_mf/1378512875Formative_Assessment_report4.pdf

Polly, D., Martin, C. S., Wang, C., Lambert, R. G., & Pugalee, D. K. (2015). Primary grades teachers' instructional decisions during online mathematics professional development activities. *Early Childhood Education Journal.* Advance online publication. DOI: 10.1007/s10643-015-0711-8

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Van den Heuvel-Panhuizen, M., Kolovou, A., & Peltenburg, M. (2011). Using ICT to improve assessment. In B. Kaur, & W.K. Yoong (Eds), *Assessment in the mathematics classroom: Yearbook 2011, Association of Mathematics Educators* (pp. 165 -185). Singapore: World Scientific and AME.

Veldhuis, M., Van den Heuvel-Panhuizen, M., Vermeulen J. A., & Eggen, T. J. H. M. (2013). Teachers' use of classroom assessment in primary school mathematics education in the Netherlands. *CADMO, 21*(2), 35-53.