

EXPLORING THE FEASIBILITY AND EFFECTIVENESS OF ASSESSMENT TECHNIQUES TO IMPROVE STUDENT LEARNING IN PRIMARY MATHEMATICS EDUCATION

Michiel Veldhuis, Marja van den Heuvel-Panhuizen

Utrecht University

In the present study we investigated to what extent workshops aimed at improving teachers' use of classroom assessment techniques had an effect on students' achievement in mathematics. Ten primary school teachers participated in two consecutive small-scale studies, aimed at using and improving different classroom assessment techniques in mathematics education. In total, 214 students were involved. The studies were carried out in the Netherlands. Qualitative and quantitative measures were used to investigate the feasibility and effectiveness of the assessment techniques. In both studies teachers and students reported enjoying the techniques and finding them useful. In terms of mathematics achievement, results indicate students improving considerably; the students' scores increased more than the national mean.

CLASSROOM ASSESSMENT TO IMPROVE STUDENT LEARNING

To gauge student learning classroom assessment by the teacher plays a pivotal role (Cizek, 2010). By using classroom assessment teachers can gather information about their students' skills and level of understanding. Collecting this information on students' learning is primordial for at least two reasons: to find out whether the instruction has had its desired effect and to generate ideas for how to proceed in the subsequent lessons. Based on assessment information teachers can align their teaching to their students' needs, which can result into adapting their teaching, but which can of course also mean not changing anything and continuing with what was planned before.

Many of the characteristics of classroom assessment appear to be part of merely good teaching practice, as Ginsburg (2009) wrote in the context of mathematics education: "Good teaching [...] sometimes involves the same activities as those comprising formative assessment: understanding the mathematics, the trajectories, the child's mind, the obstacles, and using general principles of instruction to inform the teaching of a child or a group of children (p. 126)". Nonetheless, classroom assessment can be performed in many ways, even though most teachers think of externally developed summative assessment instruments such as textbook tests or student monitoring system tests when confronted with the word assessment. Classroom assessment however is much broader than applying these instruments: it comprises all activities that permit teachers to find out where their students are at a particular moment in terms of comprehension of the subject and to give information on what is going right and wrong. Policymakers as well as influential researchers have urged the educational community, and in particular teachers, to embrace (formative) classroom assessment in

their practice. For instance, the U.S. National Council of Teachers of Mathematics (NCTM, 2013) recently took the following position on formative assessment in mathematics education: “The use of formative assessment has been shown to result in higher achievement. The National Council of Teachers of Mathematics strongly endorses the integration of formative assessment strategies into daily instruction” (p. 1).

Recently, researchers have critically examined the size of the effectiveness of (formative) assessment on student learning through reviews or meta-analyses of existing studies (e.g., Briggs, Ruiz-Primo, Furtak, Shepard, & Lin, 2012; Kingston & Nash, 2011; 2012; McMillan, Venable, & Varier, 2013). Common to these critical examinations, although their specificities differ, is that they do not contest the positive effects formative assessment is purported to have on student achievement. The matter that is under contention is the size of the effect for formative assessment on student achievement.

The present study

The purpose of the present study was to investigate the feasibility and effectiveness of classroom assessment techniques for mathematics in primary school. These classroom assessment techniques were inspired by research (e.g., Black, Harrison, Lee, Marshall & Wiliam, 2004; Torrance & Pryor, 2001), practice (e.g., Keeley & Tobey, 2011; Wiliam, 2011), and theory on classroom assessment (in mathematics education, e.g., Van den Heuvel-Panhuizen, 1996; Van den Heuvel-Panhuizen & Becker, 2003). The effectiveness of the use of separate assessment techniques (as was done for instance for the ‘Muddiest Point’ technique by Simpson-Beck, 2011) was not of interest here. Our focus was whether teachers and students were prone to use the techniques and whether the use of an ensemble of techniques would be related to an increase in achievement. Our research question was: *Do teachers like to use classroom assessment techniques (feasibility/sustainability) and is this associated with an increase in student achievement (effectiveness)?* To investigate this research question we performed two consecutive small-scale studies with groups of third-grade teachers in the Netherlands. Teachers participated in monthly workshops, consisting of three or four teachers and the first author. In these workshops classroom assessment techniques were presented, discussed, and evaluated.

METHOD

Both studies used the same method. The first part of the research question (feasibility and sustainability) was investigated by conducting regular classroom observations at least once for every teacher in between the workshops. These observations were intertwined with short informal interviews with students on their teacher’s assessment practice in mathematics. Teachers were also asked to register their evaluation of the used assessment techniques. These different sources of information are used to determine how teachers performed the classroom assessment techniques in practice, how students reacted to this, and what students and teachers alike thought of the classroom assessment techniques. To answer the second part of the research question

(effectiveness), we used a pre-/posttest evaluation of students' mathematics achievement with between the tests professional development for the teachers. The pretest data consisted of the results from the midyear student monitoring system test for Grade 3 and the results from the end of year student monitoring system test for Grade 3 (Cito LOVS; Janssen, Verhelst, Engelen, & Scheltens, 2010) served as posttest data. These tests are administered by the teachers as is common in educational practice in the Netherlands. The scores on these tests are mathematical ability scores, as calculated through item response theory models. Through the use of these test results as pre- and posttest measurement we could evaluate firstly whether the students progressed in their mathematics ability, and secondly whether students of teachers that had participated in the workshops improved more than the national sample of students of teachers that did not participate in the workshops.

Participants

Ten teachers participated in the workshops in two consecutive school years (four in Study 1 and six in Study 2). In Study 1 all teachers were female and their mean age was 38.5 years. In Study 2 two male teachers participated and the mean age was 52.5 years. In total the ten teachers taught 214 Grade 3 students (14 to 29 students per class). The teachers were found through e-mail solicitation and volunteered to participate. The schools were all situated in urbanized areas with highly mixed student populations, and the teachers used four different textbooks.

Material

We used several classroom assessment techniques in this research project. The classroom assessment techniques consisted of short activities of less than 10 minutes, which should help teachers to quickly find out something about their students, providing them with indications for further instruction, and focusing on some of the mathematics content of the second half of Grade 3. Each assessment technique was explicitly introduced as modifiable; teachers could vary the content and/or the form. This was in line with Wilson and Sloan (2000) who noted that:

[T]eachers must be: (1) Involved in the process of collecting and selecting student work. (2) Able to score and use the results immediately. (3) Able to interpret the results in instructional terms. (4) Able to have a creative role in the way that the assessment system is realized in their classrooms. (p. 191)

The forms of assessment techniques that were used in the two studies were the following: Red/Green cards, Clouds, Hard or easy, Experiment, Find the error(s), and Find problems with the same result. In Figure 1 we illustrate the Red/Green cards. Most of the techniques were centered on the assessment of number sense, mainly in the context of addition and subtraction, but the Red/Green cards could also be used to assess multiplication and division tables. In all workshops attention was paid to giving feedback to students about the assessments, so that students could become aware of their own understanding.

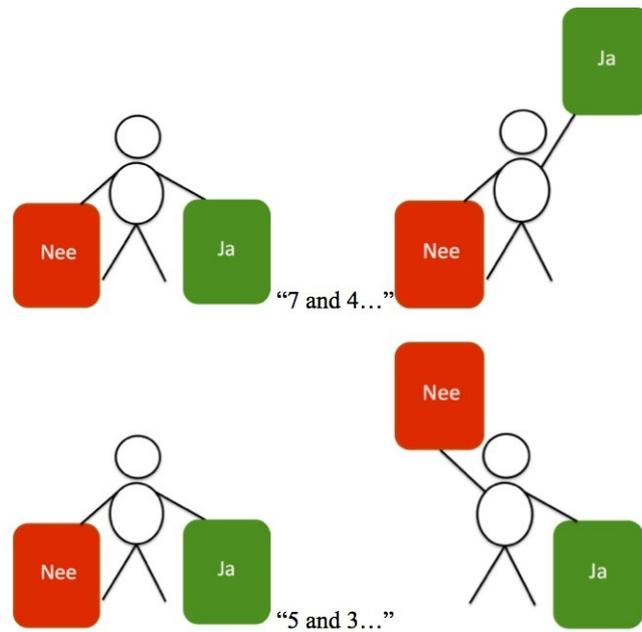


Figure 1: An example of a classroom assessment technique: The Red/Green cards.

Here teachers ask students a question that can be answered with Yes (green) or No (red). The focus is on number sense: the comprehension that two numbers together can be more or less than 10, 100, or 1000. The teacher's question in this particular example is: "Do these numbers together cross ten, yes or no?"

Procedure

Both studies had the same set-up. Teachers used several classroom assessment techniques and in doing so enlarged and reinterpreted their toolbox of assessment techniques. During the second semester of the school year the teachers and the first author convened every three to five weeks in a workshop. These workshops were organized according to the principle of 'practice what you preach' and could be considered teacher learner communities. According to Wiliam (2007) "five principles are particularly important [in establishing and sustaining teacher learning communities]: gradualism, flexibility, choice, accountability, and support (p. 197)", we strived to incorporate all of these in our workshops. As most mathematics classroom assessment techniques were embedded in or inspired by formative assessment ideas, the workshops also had a formative character. Teachers and researchers worked together in order to determine the important content in the weeks between the workshops and ways to find out whether students had learned the prerequisites or not. As such teachers "adopt and integrate these techniques and others into their own practice, they find a new synergy and see their own practice in new ways, which in turn leads to new thinking. In other words, rather than trying to transfer a researcher's thinking straight to the teacher, this new approach to formative assessment emphasizes content, *then* process (Wiliam, 2007, p. 195)". The order of business of every workshop was that first all teachers told what they had done in the preceding weeks: which assessment techniques did they use, why did they use them, in what form, how did the students react, what did they think of them, and what did they do to follow up

on what the assessment told them. These same questions were also on a feedback form the teachers were asked to fill out directly following the use of an assessment technique. When every teacher had told how their weeks had been, the researcher shared some observations made in the classrooms. The researcher visited every teacher at least one whole day between two consecutive workshops. In these visits he observed the teachers during mathematics instruction and of course the assessment techniques. As such he was able to reflect in the workshop upon what he had seen and heard in the classrooms. All the while the teachers reacted to each other's stories, they would suggest different approaches or ask for more details; generally discussion in these workshops were very lively and informative. Finally, the focus would switch to the future weeks: the content and the accompanying assessment possibilities. All would discuss these, but the researcher would after some discussion propose several ideas of which the teachers would select some and then the researcher would explain and sometimes show how the assessment technique or activity works, and in particular what could be investigated with them. Then there would be some more discussion about the activities and the researcher would present the discussed techniques on paper so that the teachers could reflect upon them in preparing their lessons.

RESULTS

An overall finding for the first part of the research question (feasibility and sustainability) from the classroom observations, the interviews, and the discussions in the workshops was that every teacher, even though they participated in the same workshop and got the same assessment techniques to work with, interpreted the classroom assessment in their own way and adapted them to their own practice. For instance, the Red/Green cards technique seems quite straightforward on paper; nonetheless there was great variation in how teachers performed this technique in their respective classrooms. A teacher of Study 1 noticed that some students waited to see which card other students held up before choosing their own. She considered this a problem "as it was a testing situation" and decided that students had to be in "testing positions" (separated tables) and even close their eyes so that they could not cheat. Another teacher of Study 1 spend quite some time to ensure that all students were clear about what the colours green and red were, and subsequently in which hand they held each colour. A teacher of Study 2 interpreted the Red/Green cards more as a game, and adapted it to his own practice. He considered it to be "nonsense to be the only one doing the work" and let a student (every time a different one) come up with the problems on the spot. These three short examples show how diversely teachers operated in their classrooms and how flexibly they used this assessment technique. For the second part of the research question (effectiveness) we compared the pre- and posttest data of the student monitoring system tests for every study separately.

Study 1

For the first study, as can be seen in Table 1, the mean ability of students increased from midyear to end of year testing.

	Class	<i>N</i>	Mean ability midyear – end of year	Ability gain	Effect size (<i>d</i>)
Study 1	I	15	65.3 – 76.7	11.4	1.30
	II	13	62.2 – 68.4	6.2	0.42
	III	14	71.6 – 83.2	11.6	0.73
	IV	24	75.8 – 85.3	9.5	0.78
	<i>Total</i>	<i>66</i>	<i>68.7 – 78.4</i>	<i>9.7</i>	<i>0.81</i>
National mean			69.0 – 74.1	5.1	0.36

Table 1: Mean ability, gain score, and effect size per class for the two studies, and the Dutch national mean on the Cito LOVS tests.

It was to be expected that students progress in their mathematical ability whether they have teachers that perform specific assessment activities or not, just as a result of growing older and having more education; the national mean also shows this direction. However, the mean difference between pre- and posttest over the four classes of participating teachers and the effect size (gain score = 9.7 ability points, $d = 0.81$) are notably larger than those of the national norm sample (gain score = 5.1, $d = 0.36$). This means that students of teachers in Study 1 progressed 0.45 SD more in three months' time than students in the national sample.

Study 2

For the second study we had not one workshop, but two separate workshops in two different cities. These workshops were of a slightly different nature than in Study 1, for one the ages of the participating teachers were quite different, also the frequency of meetings was down from every three weeks to once per month or even every five weeks, and finally the assessment techniques were provided in a slightly more definite way. The mean ability gains from midyear to end of the year are displayed in Table 2.

	Class	<i>N</i>	Mean ability midyear – end of year	Ability gain	Effect size (<i>d</i>)
Study 2	V	22	66.6 – 72.2	5.6	0.32
	VI	17	66.3 – 76.1	7.8	0.44
	VII	26	67.9 – 73.1	5.1	0.51
	VIII	26	72.8 – 82.6	9.8	0.67
	IX	28	74.8 – 83.6	8.9	0.74
	X	27	75.6 – 83.8	8.2	0.61
	<i>Total</i>	<i>149</i>	<i>71.0 – 78.6</i>	<i>7.6</i>	<i>0.55</i>
National mean			69.0 – 74.1	5.1	0.36

Table 2: Mean ability, gain score, and effect size per class for the two studies, and the Dutch national mean on the Cito LOVS tests.

Just as for Study 1 we observe that all classes improved on average, and that with an ability gain of 7.6 and effect size of $d = 0.55$. This score gain was bigger than the one in the national sample, which was of 5.1 ability points with an effect size of $d = 0.36$. This means that students of teachers in the second study progressed 0.19 SD more in three months' time than students in the national sample.

DISCUSSION

The feasibility of the classroom assessment techniques combined with an indication that they were effective are the main results of this study. The gain scores of the two small-scale studies showed that students learned considerably more when teachers make effective use of classroom assessment, than students from the national sample. This relative gain was quite large (between 0.19 and 0.45 standard deviation), while the professional development –the intervention– only took between four and six meetings of about an hour. However, given the fact that there was no control group the direct attribution of these learning effects to the sole use of classroom assessment techniques would be too simple.

The second study was performed to fine-tune the techniques as well as to find out whether a lower frequency of workshops and researcher visits would still be associated with a learning effect. In this second study we found approximately the same results as in Study 1: a larger than normal learning effect for the students of participating teachers. As such this provided supplementary evidence for the effectiveness of the techniques; however the same problems as in Study 1 persisted. The difference in Study 2 was slightly smaller than in Study 1. Several things could explain this difference in gains, for one there were less contact moments with other teachers and the researcher, which could have led to less implication in the research in the second study. This came forward in one of the workshops in Study 2, where some teachers admitted that they had only performed the classroom assessment techniques the morning preceding the workshop. The urgency and enthusiasm to use the techniques that they had voiced in the previous workshop had quickly diminished after it had ended. In Study 1 teachers used the techniques generally in the week following the workshop and often repeatedly until the next meeting. Quite understandably this could have caused a differential effect on student learning. An additional explanation as to why some of the teachers in the second study seemed less invested could be their ages. The teachers in Study 2 were older and had as such more experience in teaching than teachers in Study 1. Some of these older teachers did not believe in all the techniques and were less inclined to use some of them, whereas this did not occur with the slightly younger group in Study 1.

These considerations can be taken into account in the design of further research on the effectiveness and the use of classroom assessment techniques.

References

- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in education: Principles, Policy & Practice*, 5(1), 7-74.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E. M., Shepard, L. A., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.

- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). Abingdon, UK: Routledge.
- Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, 52, 109-128.
- Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8* [Scientific justification of the mathematics test for Grade 1 until Grade 6]. Arnhem, Netherlands: Cito.
- Keeley, P., & Tobey, C. R. (2011). *Mathematics formative assessment: 75 practical strategies for linking assessment, instruction, and learning*. Thousand Oaks, CA: Corwin.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.
- Kingston, N., & Nash, B. (2012). How many formative assessment angels can dance on the head of a meta-analytic pin: 2. *Educational Measurement: Issues and Practice*, 31(4), 18-19.
- McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: so much more is needed. *Practical Assessment, Research & Evaluation*, 18(2). Retrieved from <http://pareonline.net/getvn.asp?v=18&n=2>.
- National Council of Teachers of Mathematics. (2013). *Formative assessment: A position of the National Council of Teachers of Mathematics*. Retrieved from http://www.nctm.org/uploadedFiles/About_NCTM/Position_Statements/Formative%20Assessment1.pdf
- Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631.
- Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Universiteit Utrecht, Utrecht.
- Van den Heuvel-Panhuizen, M., & Becker, J. (2003). Towards a didactical model for assessment design in mathematics education. In A. J. Bishop, M. A., Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 689-716). Dordrecht: Kluwer Academic Publishers.
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181-208.