Using classroom assessment techniques in Chinese primary schools: Effects on student achievement

Xiaoyan Zhao, Marja van den Heuvel-Panhuizen, Michiel Veldhuis

Using classroom assessment techniques in Chinese primary schools: Effects on student achievement

Xiaoyan Zhao[1], Marja van den Heuvel-Panhuizen[1,2] and Michiel Veldhuis[2]

[1]Freudenthal Institute, FBW, Utrecht University, the Netherlands; x.zhao@uu.nl

[2]Freudenthal Group, FSW, Utrecht University, the Netherlands

In an experimental study with a pretest/posttest/delayed posttest and control-group design, we investigated the effects on students' mathematics achievement of using classroom assessment techniques in Chines classrooms. Participants were 47 third-grade teachers and their 608 students in Nanjing, China. The teachers were assigned to either the experimental condition, participating in two two-hour workshops on classroom assessment, or the control condition, in which the teachers followed their regular teaching plans. The workshops focused on the use of classroom assessment techniques to reveal students' understanding of multiplication and to enable teachers to adapt teaching to their students' needs. Students from the teachers in the experimental condition slightly improved their mathematics achievement scores. However, no statistically significant difference was found between the two conditions.

Keywords: Classroom assessment, student achievement, China, multiplication, teachers.

Introduction

The guidance teachers provide in their mathematics classes to their students can be more or less effective for stimulating students' learning processes, depending on whether their instruction is attuned to students' needs and possibilities for further development. Therefore, at practically every moment teachers need to know where the students are in their learning process (Wiliam, 2011). This was also recently emphasised by Schoenfeld (2014) when he wrote that "[p]owerful instruction 'meets students where they are' and gives them opportunities to move forward" (p. 407). Classroom assessment, i.e. assessment in the hands of the teachers that is interwoven with instruction and integrated in daily teaching practice, can inform teachers of 'where their students are' and as such enable them to adapt their further instruction to their students' needs.

Since the importance of classroom assessment on raising students' achievement was revealed by Black and Wiliam (1998), much attention has been paid to professional development to enhance teachers' classroom assessment practice. The rationale for this is that providing professional development to teachers on the use of classroom assessment can lead to teachers gaining more information on their students' understanding and skills. Through this information teachers can adapt their teaching to their students' needs, which in turn is expected to lead to improved student achievement. Whether professional development indeed has impact on student achievement was investigated in several studies (Phelan et al, 2012; Randel, Apthorp, Beesley, Clark, & Wang, 2016; Thompson, Paek, Goe, & Ponte, 2004; Veldhuis & Van den Heuvel-Panhuizen, 2014, 2016). The results of these studies are mixed. Facilitating teachers to use classroom assessment has been shown to lead to considerable improvement of students' achievement (Phelan et al, 2012, Veldhuis & Van den Heuvel-Panhuizen, 2014, 2016). It also happened that professional development on classroom assessment had only a small but consistent positive effect on student learning (Thompson et al, 2004) or failed to yield any statistically significant impact (Randel et al, 2016). These mixed, but generally positive, results on the effects of professional development on classroom assessment were all found in the western educational context. As there are important differences between mathematics education in Western and East Asian countries (Leung, Graf, & Lopez-Real, 2006), we aimed to find out whether giving support to Chinese teachers on the use of classroom assessment would have an effect on their students' mathematics achievement.

In China, recently, classroom assessment has received increasing attention from primary school mathematics teachers, as evidenced by an increasing number of teacher-written papers addressing classroom assessment (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2017). Moreover, in Chinese primary mathematics education, teachers generally agree that assessment is useful for the improvement of teaching and learning, and they assess their students at least weekly by employing various methods, for example observing, questioning and assigning textbook tests (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2016a). It seems that classroom assessment has been widely embraced and implemented in teaching practice. However, professional development focused on classroom assessment seems sparse (Zhao et al, 2016a), let alone investigations into its effect on students achievement.

In our study, classroom assessment is conceived as the use of what we call 'classroom assessment techniques' (CATs): short teacher-initiated assessment activities that teachers can use in their daily practice to reveal their students' understanding of a particular mathematical concept or skill. These CATs have been used in earlier research in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014, 2016). Our main research question was: *What are the effects of supporting Chinese primary school mathematics teachers' use of classroom assessment techniques (CATs) on students' mathematics achievement?*

Method

An experiment with pretest/posttest/delayed posttest and control-group design (see Table 1) with 47 third-grade mathematics teachers from 18 primary schools was carried out in Nanjing, China. All teachers used the same textbook, namely 苏教版 textbook published by Jiangsu Phoenix Education Publishing House (2014). Based on the participating schools' reputation, educational quality, and location, pairs of matched schools were allocated either to the control or to the experimental condition. Teachers in the experimental group participated in two two-hour workshops on the use of classroom assessment techniques, whereas the teachers in the control group followed their regular teaching plans.

Condition	January	March			May
		Week 1	Week 2	Week 3	
Control	Pretest			Posttest	Delayed posttest
Experimental	Pretest	Workshop	Workshop	Posttest	Delayed posttest

Table 1: Time schedule of the experiment in 2015

In the workshops, the teachers were introduced to eight CATs. These CATs are low-tech and low-cost, and can easily be implemented by teachers. Every technique consists of a short activity (less than 10 minutes) and helps teachers to quickly find out something about their students' understanding of mathematics, provides indications for further teaching. Also, the teachers could adapt the techniques to their own practice; they could choose when and how to use the CATs. The focus of the assessment techniques was on the first chapter of the second semester of Grade 3, in which students learn how to solve multiplication problems of two-digit numbers mainly by written digit-based algorithm. In the following we illustrate three examples of these CATs. During the workshops, the teachers in the experimental condition were provided with a detailed teacher guide describing the eight CATs that all fitted to the content of their textbook. Detailed information about the purpose of the CATs and suggestions for how to use them was provided and discussed during these workshops. It was also explained that the teachers were free to decide how they would use the CATs in practice in the following two weeks of multiplication teaching.

CAT 1: Family problems

This CAT (see Figure 1) is aimed at assessing whether students recognize similarities among analogous problems and can use the given answer to one of these problems to solve the others.

It is known that **97×8** equals **776**.
Do you think you can solve the following problems?
(Yes- Green card ; No- Red card)

a) 97×80 b) 97×800 c) 97×8000 d) 970×8000

Figure 1: CAT 1: Family problems

One strategy to solve a multiplication problem with either the multiplicand or the multiplier being a multiple of 10 is making use of an analogous problem of which the answer is known or which is easy to calculate. A requirement for students to choose and use this strategy is that they understand the analogous relationship, even when the numbers involved in the multiplication are bigger than two digits. CAT 1 is meant to elicit information of whether and to what extent students have this understanding. The students are provided with the answer of 97×8 and are then asked whether they *think* they are able to solve mentally a number of other, related multiplication problems that, at first sight, are not easy to solve by mental calculation. CAT 1 differs from the regular assessment tasks in the textbook in which the students have to carry out the calculation and the focus is on detecting whether students can do this correctly. In CAT 1 it is assessed whether the students recognize the analogue structure of the problems and are aware that they can use this for solving these problems. In CAT 1, the teacher asks for every problem whether students think they are able to solve it. All students have a green card (for the answer: "Yes") and a red card (for the answer: "No") with which they can show their answers (see Figure 2). By inspecting the waving green and red cards the teacher gets an immediate overview of the students' responses and whether they see the analogy between the problems, and whether their understanding is affected by the number of zeroes in the family problems.

Figure 2: Students showing their cards in CAT 1 for the problems 97×80 and 970×8000

CAT 2: Breaking down a multiplication

This CAT (see Figure 3) is aimed at assessing whether students can identify the components of a multiplication by filling in the blanks on a work sheet.

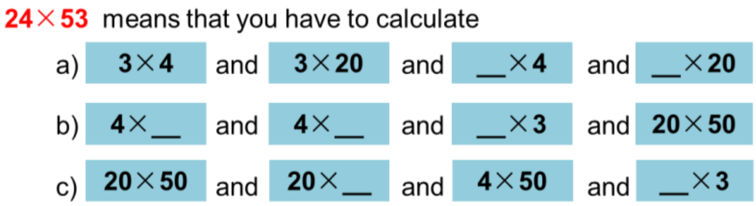

Figure 3: CAT 2: Breaking down a multiplication

Students may be able to find the correct answer of a problem like 24×53 by performing the standard multiplication algorithm perfectly; however, this does not necessarily mean that students understand what they are doing and that they understand the structure of multiplications with multi-digit numbers, which is the focus of CAT 2. This approach of requiring students to unravel multiplication problems differs from the regular approach to assessing students in which finding the correct answer of a multiplication problem receives most attention of mathematics teachers. In the case of CAT 2, the multiplication of 24 and 53 can be unpacked into four sub-multiplications, namely 3×4, 3×20, 50×4, and 50×20. The sum of the results of these sub-multiplications gives the answer of 24×53. By asking students to identify the components of a multiplication problem of multi-digit numbers it can be revealed whether they understand what is 'behind' the multiplication algorithm. For example, the student work in Figure 4 shows that Student 1 has difficulties in being fully aware of the values of the digits (having 5×4 and 5×20 instead of 50×4 and 50×20 in Task a, and having 2×3 instead of 20×3 in Task b), while Student 2 could not clearly distinguish the different components of the multiplication 24×53 (having 4×3 instead of 20×3 in Task b, and having no answer filled in Task c).

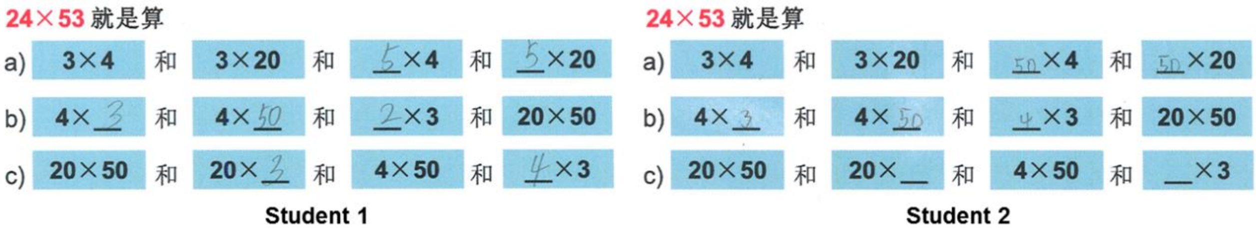

Figure 4: Work of two students in CAT 2

CAT 3: Fruit language

This CAT (see Figure 5) is aimed at assessing whether students can use the associative and distributive property of multiplication to restructure a multiplication problem.

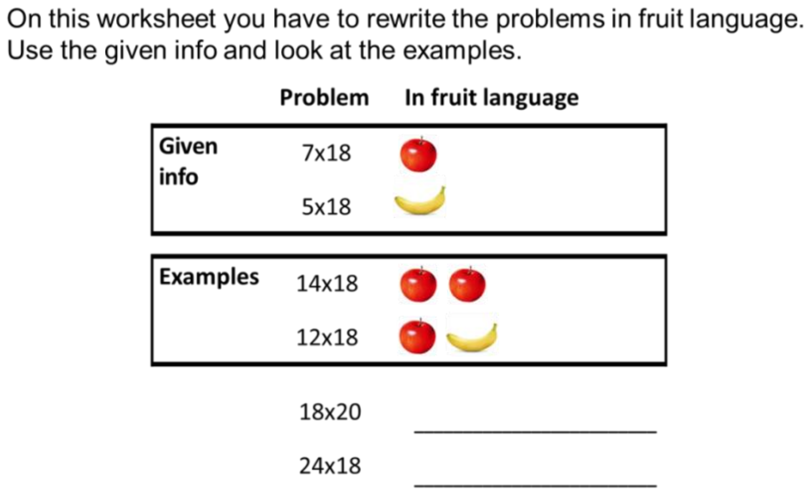

Figure 5: CAT 3: Fruit language

Making use of the associative and distributive property of multiplication is the basis of solving multiplication problems. By using these properties students can convert a difficult multiplication problem into a number of easier multiplication problems. For example, 25×36 can be solved by calculating 20×36 and 5×36 (distributive property) or by calculating 25×4×9 (associative property). For solving multiplication problems in this way it is very important that students understand the associative and distributive property of multiplication and that they can identify the possibilities of restructuring a multiplication problem. CAT 3 provides an opportunity for students to show this understanding. In order to avoid the difficulty of formal notations, fruit is used as a substitute.

The student work shown in Figure 6 reveals that Student 3 has arrived at a high level of the understanding of the associative and distributive property of multiplication and is able to notate this in a proper mathematical way, although not using a formal notation with number or letter symbols. Student 4 only 'rewrote' one of the multiplication problems (18×20) by drawing four bananas. Moreover, the worksheet of this student shows that he/she did not use the properties of multiplication but instead was calculating the multiplications and then tried to express the answer by using the fruit.

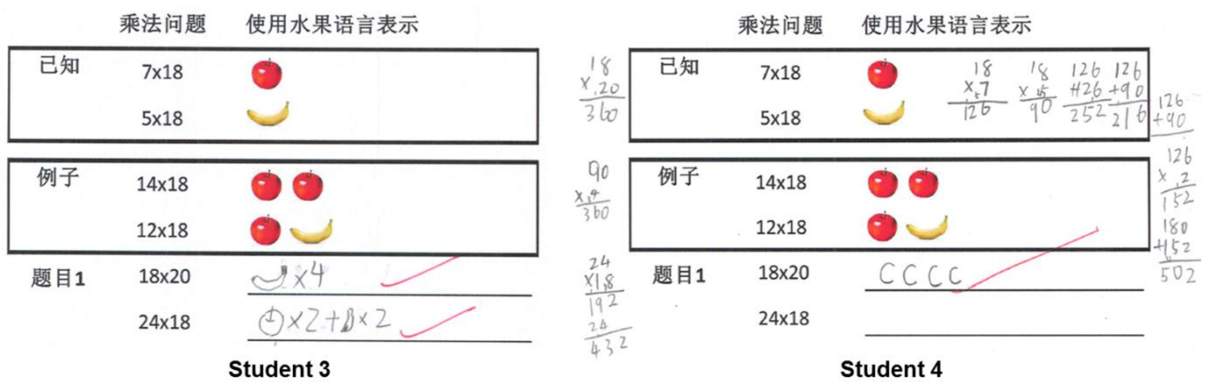

Figure 6: Work of two students in CAT 3

In order to measure students' mathematics achievement, three tests were used, which were designed and arranged by the local teaching research office. These tests have the same structure in terms of the type of questions and total score (100 points). However, the mathematical domains that are tested are different. The immediate posttest was an end-of-chapter test and focused on the multiplication of two-digit numbers. The pretest and the delayed posttest were end-term and mid-term tests, which also included problems related to measurement, fractions, and geometry. Nevertheless, multiplication is the main focus of all the three tests (30% of the points in the pretest and delayed posttest and 90% in the immediate posttest were related to multiplication tasks).

Originally, 3040 students took the tests. Since it was found that mistakes were made when grading students' examination papers, we decided to choose 608 (20%) students systematically, based on their student number in every class, for data checking to be included in the final analysis.

Results

Unexpectedly, on average, students in both conditions had decreasing mathematics achievement scores from pretest (M_{exp} = 89.2, SD_{exp} = 8.7; M_{con} = 90.8, SD_{con} = 7.7) to immediate posttest (M_{exp} = 88.5, SD_{exp} = 9.3; M_{con} = 89.5, SD_{con} = 9.0) and to delayed posttest (M_{exp} = 86.4, SD_{exp} = 12.2; M_{con} = 87.7, SD_{con} = 11.2). When looking at the standardized scores this image becomes a bit less clouded by the different tests measuring different domains at the different time points, therefore we report the *z*-scores in Table 2. The pattern remains almost the same, with relatively higher scores in the control condition than in the experimental condition, but, in the experimental condition, a slight improvement of the scores appears after the intervention.

	Pretest score		Posttest score		Delayed posttest score		
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*n*
Control	0.104	0.930	0.059	0.986	0.058	0.952	278
Experimental	-0.088	1.049	-0.050	1.010	-0.049	1.038	330

Table 2: Descriptive statistics of students' standardized mathematics (z) scores per condition for the pretest, posttest, and the delayed posttest

We performed an analysis of covariance (ANCOVA) on the immediate posttest scores to see if this small improvement was statistically significant. In this ANCOVA the pretest score was entered as covariate and condition as fixed factor. It turned out that no significant effect for condition was found ($F(1, 605) = 0.08$, $p = .776$, $\eta_p^2 = 0.000$).

Discussion

The students of the teachers that participated in the workshops on the CATs only very slightly improved their standardized mathematics achievement scores after the intervention. This improvement was not significant, neither in size, nor in the statistical sense. Contrary to these findings in the experimental group, the students in the control group did not improve their standardized scores from one test to the other. However, on average the students in the control condition outperformed the students in the experimental condition on all three tests. A possible reason for the minor changes in students' mathematics achievement could be that there appeared to be a strong ceiling effect on the tests (average success scores of around 90%). Maybe students' extant high achievement level could also have caused that the use of the CATs did not further optimize the teachers' instruction. Another explanation for the small improvement in the experimental condition could be the short period of time of the intervention. In less than three weeks, the teachers in the experimental condition needed to understand how to use the CATs, to incorporate them into their teaching plans, and to reconcile the new insights into their students with their original understanding of students and teaching. For teachers to really get used to and to make the most of the implementation of the CATs, probably more time needs to be reserved and more guidance needs to be offered in the professional development workshops.

Also the context of the experimental study may have influenced the effect of the CATs on the students' mathematics achievement. First of all, as we found in an earlier study (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2016b) Chinese primary school mathematics teachers have detailed lesson plans and tend to include CATs in their pre-arranged lessons as extra exercises rather than implementing them as formative assessment activities. As such, the teachers may not have used the information gathered with the CATs for adapting their instruction. Another issue is that the planned lessons have for every addressed topic a fixed time schedule for instruction and practice. By including the CATs less time could be spent on teaching these topics and students may have had less practice in solving the problems as used in the regular tests. A promising finding is that despite this smaller investment in the regular program the students in the experimental condition did not perform really worse in the regular tests than their counterparts in the control condition. In this way, our study provides some evidence which may encourage teachers to go beyond the straightforward testing of the standard operations and pay also attention to examining students' deeper understanding of these operations, and use the assessment information adaptively for improving instruction and student learning.

Acknowledgments

This work was supported by the China Scholarship Council (CSC) under Grant 201206860002; and the Netherlands Organization for Scientific Research (NWO) under Grant NWO MaGW/PROO: Project 411-10-750. The authors thank Prof. Lianhua Ning of Nanjing Normal University, China, for helping to contact schools and all the teachers involved in this study for their cooperation and contribution.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Jiangsu Phoenix Education Publishing House. (2014). *苏教版教科书(小学数学三年级下册) [苏教版 Textbook (Mathematics textbook for Grade 3 in primary education, Volume 2)]*. Nanjing: Author.

Leung, F. K. S., Graf, K. D., & Lopez-Real, F. J. (2006). Mathematics education in different cultural traditions: a comparative study of East Asia and the West. In F. K. S. Leung, K. D. Graf, & F. J. Lopez-Real (Eds.), *Mathematics Education in Different Cultural Traditions. A Comparative Study of East Asia and the West. The 13th ICMI Study* (pp. 1–20). New York: Springer.

Phelan, J. C., Choi, K., Niemi, D., Vendlinski, T. P., Baker, E. L., & Herman, J. (2012). The effects of POWERSOURCE© assessments on middle-school students' math performance. *Assessment in Education: Principles, Policy & Practice, 19*(2), 211-230.

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research, 109*(5), 491-502.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher, 43*(8), 404-412.

Thompson, M., Paek, P., Goe, L., & Ponte, E. (2004). *Study of the California formative assessment and support system for teachers: Relationship of BTSA/CFASST and student achievement*. Retrieved from https://www.ets.org/Media/Research/pdf/RR-04-32.pdf

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014). Exploring the feasibility and effectiveness of assessment techniques to improve student learning in primary mathematics education. In C. Nicol, S. Oesterle, P. Liljedahl & D. Allan (Eds.), *Proceedings of PME38 and PME-NA36, Vol. 5* (pp. 329–336). Vancouver, BC: PME.

Veldhuis, M. & Van den Heuvel-Panhuizen, M. (2016). *Supporting primary school teachers' classroom assessment in mathematics education: Effects on students' learning.* Manuscript submitted for publication.

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

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2016a). *Chinese primary mathematics teachers' views on assessment: Findings from a large-scale questionnaire survey*. Manuscript in preparation.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2016b). Teachers' use of classroom assessment techniques in primary mathematics education – An explorative study with six Chinese teachers. *International Journal of STEM Education,* 3: 19. doi: 10.1186/s40594-016-0051-2

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2017). Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers. *Studies in Educational Evaluation, 52,* 42-54.