

## Assessment in the Hands of Primary School Mathematics Teachers: Four studies in China

Xiaoyan Zhao<sup>1,2</sup>   
 Marja van den Heuvel-Panhuizen<sup>1,3,4</sup>   
 Michiel Veldhuis<sup>3,5</sup> 

<sup>1</sup> Utrecht University, Faculty of Science, Freudenthal Institute, Utrecht, Utrecht, Nederland

<sup>2</sup> School of Teacher Education, Nanjing Normal University, China

<sup>3</sup> Utrecht University, Faculty of Social and Behavioural Sciences, Freudenthal Group, Utrecht, Utrecht, Nederland

<sup>4</sup> Nord universitet, Fakultet for lærerutd., kunst og kultur, Bodø, Nordland, Norge

<sup>5</sup> Hogeschool iPabo, Amsterdam, Noord-Holland, Nederland

*Received for publication on 28 Mar. 2019. Accepted, after revision, on 15 Apr. 2019.*

### ABSTRACT

In this paper, we describe four studies that have been done in the project Improving Classroom Assessment in China (ICA-C). The project was a sequel to the ICA project in the Netherlands. To shed light on the current situation of Chinese primary school mathematics teachers' perception and practice of classroom assessment, we first did a review study based on teacher-written papers addressing classroom assessment. Then a large-scale questionnaire survey was conducted by means of which we could identify different assessment profiles of teachers. With the gained knowledge, two other studies were set up to explore the possibility for improving Chinese primary school mathematics teachers' assessment activities, with a particular focus on using classroom assessment techniques in the domain of division and multiplication.

**Keywords:** Classroom assessment, Mathematics Education, Primary School, China.

## Avaliação nas Mãos de Professores de Matemática da Escola Primária: Quatro Estudos na China

### RESUMO

Neste artigo, descrevemos quatro estudos que foram feitos no projeto Improving Classroom Assessment in China (ICA-C). O projeto foi uma continuação do projeto ICA na Holanda. Para esclarecer a situação atual da percepção dos professores de matemática da escola primária chinesa e da prática de avaliação em sala de aula, primeiro fizemos um estudo de revisão com base em artigos escritos por professores que abordavam a avaliação em sala de aula. Em seguida, realizou-se uma pesquisa por questionário em grande escala, por meio da qual pudemos identificar diferentes perfis de avaliação dos professores. Com o conhecimento adquirido, dois outros estudos foram criados para explorar a possibilidade de melhorar as atividades de avaliação dos professores de matemática das escolas primárias chinesas, com um foco particular na utilização de técnicas de avaliação em sala de aula no domínio da divisão e multiplicação.

**Palavras-chave:** avaliação em sala de aula, educação matemática, escola primária, China.

Corresponding author: Xiaoyan Zhao. E-mail: x.zhao@uu.nl

Acta Scientiae	Canoas	v.21	n.2	p.128-136	mar./abr. 2019
----------------	--------	------	-----	-----------	----------------

## **CLASSROOM ASSESSMENT IN MATHEMATICS EDUCATION**

Increasing the integration of teaching, learning and assessment has been identified as one of the key trends in the field of assessment in mathematics education since the last century (Ruthven, 1994). Awareness has arisen that assessment should not only serve summative purposes for ranking and selecting students but should place more emphasis on formative purposes for informing teachers' instruction and improving students' learning (Black & Wiliam, 1998). Also, contrary to the low reliability that, in the past, was attributed to teachers' judgements of students' performance (Parkes, 2013), nowadays, teachers' crucial role in their assessment of students has been highly valued (Cizek, 2010). Formative assessment activities in the hands of teachers are referred to as classroom assessment (e.g., De Lange, 1999). These teacher-led assessment activities that are interwoven with instruction and fully integrated in the teachers' daily teaching practice, such as questioning, observing students, and giving quizzes or teacher-made written assignments, can provide instructional insights about students' thinking and about what productive and actionable next steps might be taken in order to meet students' needs (Shepard, Penuel, & Pellegrino, 2017). In this way, classroom assessment entails the potential to move students' learning forward. In fact, promoting the use of classroom assessment in mathematics education has been put on the worldwide policy agendas (e.g., NCTM, 2013; CDC, 2002). In line with this, studies have been carried out with the focus on investigating and improving primary school teachers' classroom assessment in mathematics education (e.g., Andersson & Palm, 2017; Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2014b).

## **A RESEARCH PROJECT ON CLASSROOM ASSESSMENT IN CHINA**

In mainland China, where there exists a deeply-rooted examination culture focusing on ranking and selection, in 2001, a new approach to assessment was launched by the Ministry of Education (MoE, 2001) aimed at supporting teaching and learning. Since then, only a few studies provided information about the implementation of this assessment reform in primary school mathematics education (Brown, Hui, Yu, & Kennedy, 2011; Ni, Li, Li, & Zou, 2011; Zhao, Mulligan, & Mitchelmore, 2006). The most recent data collected by these studies date from 2008, which means that little is known about how the implementation of the assessment reform has further evolved. Therefore, in 2012, the Improving Classroom Assessment in China (ICA-C) project was started. This project was a sequel to the ICA project in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2019) and contained four studies. *Study 1* was a review study in which existing resources of journal papers addressing classroom assessment written by teachers were examined. *Study 2* was a large-scale survey in which an adapted version of the questionnaire developed in the Dutch ICA project was used to collect data for identifying teachers' assessment profiles. After having acquired some understanding of how Chinese primary school mathematics teachers consider and perform classroom assessment, two other

studies were set up to explore the possibility for improving their assessment activities, with a particular focus on using classroom assessment techniques. Classroom assessment techniques (CATs) are short teacher-initiated targeted assessment activities proximate to the textbook, which teachers can use in their daily practice to reveal their students' understanding of a particular mathematical concept or skill. *Study 3* was an explorative study to investigate how Chinese primary school mathematics teachers used CATs that were meant to assess students' understanding of division. *Study 4* was an intervention study aimed at examining what insights teachers can gain from using CATs into their students' mathematical understanding in the domain of multiplication.

## **Study 1: Classroom Assessment in the Eyes of Teachers**

In a review study (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2017), 266 teacher-written papers, which were addressing classroom assessment and published in 2011 and 2012, ten years after the start of the assessment reform, were identified in the China National Knowledge Infrastructure database to answer the research question: *What do teacher-written papers reveal about Chinese primary school mathematics teachers' conceptions of classroom assessment and how these conceptions are related to the assessment guidelines as included in the mathematics curriculum standards released in 2011?* For analysing these papers, a coding framework was developed, of which the main categories were in line with the main aspects of assessment provided in the assessment guidelines in the mathematics curriculum standards (MoE, 2011). This included the purpose of assessment, the content of the assessment, the person who is the assessor, the method of assessment, the provision of feedback, and using assessment results for instructional adaptation. When deciding about further subcategories, the themes the teacher-authors brought up were also taken into account.

The results revealed that the teacher-authors reflected often on the purpose of assessment (49% of the 266 papers), the content to be assessed (70%), the person who is the assessor (100%), and the used assessment methods (78%). Particularly, providing feedback was addressed very frequently. In 198 papers (74% of the 266 papers), at least a quarter of the text referred to feedback, in 64 of these, the teacher-authors considered classroom assessment even to be equivalent to feedback. Another finding was that in only 9 papers (3%) there was evidence that the teacher-authors used the gathered assessment information for instructional adaptation. In general, it was found that the teachers' conceptions of classroom assessment, as revealed in the teacher-written papers, echoed quite well with nearly all aspects of the assessment guidelines, except for using assessment results for instructional adaptation.

## **Study 2: Teachers' Assessment Profiles**

A large-scale survey (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2018) was conducted to answer the research question: *What assessment profiles can be identified*

*in Chinese primary school mathematics teachers?* For collecting data, an existing questionnaire developed in the ICA project in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014a; Veldhuis, Van den Heuvel-Panhuizen, Vermeulen, & Eggen, 2013) was adjusted to fit the Chinese context. The adapted questionnaire consisted of 30 questions to gather information about teachers' background, their general teaching practice, and their assessment practice and beliefs. In the end, the responses of 1101 Chinese primary school mathematics teachers from 12 provinces and regions were analysed. Through exploratory factor analyses, the underlying structure of the questionnaire was uncovered and eight factors were determined: (1) general instructional decision-making assessment purposes, (2) specific instructional decision-making assessment purposes,

(3) assessment methods, (4) diversity of assessment problem format, (5) importance of assessing skills and knowledge, (6) importance of assessing extra-curricular skills, (7) perceived usefulness, and (8) acceptance of assessment. With a latent class analysis, three assessment profiles were identified. When looking into the assessment profiles through the lenses of the eight factors, distinct characteristics regarding teachers' views on assessment became clear.

The largest group of teachers belonged to the Mainstream assessors' profile (53.1% of the teachers). These teachers appeared to be moderate in their use of assessment, as they scored quite close to the mean on most factors and relatively high on acceptance of assessment. In detail, these teachers reported using several assessment methods for different purposes of instructional decision-making with an average frequency. To assess students, they reported using several different problem formats. In addition, these teachers generally underlined the importance of assessing different types of skills and knowledge and acknowledged assessment to be useful for supporting teaching and learning.

Moreover, these teachers were, among the teachers in the three assessment profiles, most acceptant of using assessment in their practice. The second group of teachers belonged to the Enthusiastic assessors' profile (21.7%), who had above average scores overall. They reported using different assessment methods very frequently for various purposes, highly endorsed the importance of assessing different skills and knowledge and perceived assessment to be very useful. The remaining teachers (25.2%) were considered as Unenthusiastic assessors. These teachers scored on almost all factors far below the mean, indicating that they did not report to use assessment purposefully or regularly, and did not deem it to be essential or useful.

### **Study 3: Teachers' Use of CATs**

In an explorative study (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2016), Chinese primary school mathematics teachers' use of CATs was investigated. Six female third-grade mathematics teachers from two primary schools in Nanjing were offered a series of CATs that were designed to assess their students' understanding of the division of three-digit numbers by a one-digit number. For developing these CATs, we followed the same process as that had been done in the Dutch ICA project. Textbooks analysis was

the starting point to make assessment close to teachers' daily teaching. Then, decisions needed to be made on the basis of mathematical-didactical analyses about what content to be assessed, and about what questions should be asked in order to give teachers access to students' deep understanding. Crucial for this is that the questions take a different perspective than that offered in the textbook. Finally, the CATs should have a format that makes them be used feasibly (Veldhuis & Van den Heuvel-Panhuizen, 2014b, 2019). Here we provide one CAT as an example (Figure 1). In CAT-1 identifying the watershed, after showing the task, the teacher mentions the possible divisors one by one from 1 to 9. For each divisor, students need to indicate the digits of the quotient by raising either green or red card. This CAT, with the format of red/green cards, is meant to quickly give the teacher an overview about whether students can identify the breaking point (the watershed) of the possible divisors when the number of digits in the quotient changes.

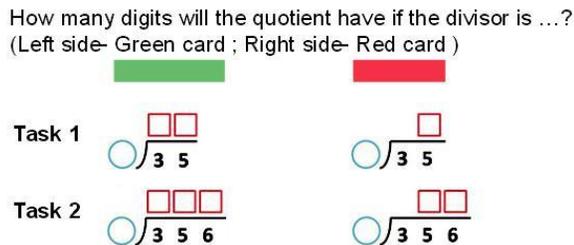


Figure 1. CAT-1 identify the watershed, a CAT with the format of red/green cards for assessing students' understanding in the domain of division.

To help the teachers understand and use the CATs, a teacher guide was developed in which the purpose of each CAT and the suggestions for its implementation were described. Moreover, four one-hour meetings were organized in which how to use CATs was discussed and teacher' experience with CATs was shared. Data regarding teachers' use of CATs were collected by teacher interviews, feedback forms, and final reports. Additionally, data from lesson observation and student work were used, if necessary, to mirror or supplement the information reported by the teachers.

It was found that, although CATs were quite new to the teachers, they could easily include CATs in their practice by changing them to fit their pre-arranged lesson plans. By using the CATs, the teachers got new information about their students' mathematics understanding. They valued the way in which CATs challenged their students with questions that were not completely prepared by textbooks, and thus CATs were very revealing for them. In addition, most teachers liked the CATs with the format of the red/green cards for providing quick information about students' understanding. However, no evidence was found that the teachers used the information gained from the CATs for adapting their instruction in the subsequent lessons to meet the students' needs. Instead of adapting their further teaching, the teachers gave their students, during or after carrying out the CATs, instant help for arriving at correct answers. In addition, the teachers even used the CATs to adapt their instruction beforehand.

## Study 4: Insights Teachers Gained from Using CATs

The intervention study (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2019) that followed aimed to examine what insights Chinese primary school mathematics teachers gained into their students' mathematical understanding from using CATs. Its setup was similar to that in the explorative study (*Study 3*); however, a larger group of teachers was involved, and a different mathematics topic was chosen. In total, 25 third-grade mathematics teachers from nine schools in Nanjing were given eight CATs assessing students' understanding of multiplication of two-digit numbers. One example of these CATs is provided (Figure 2). In CAT-2, students need to break down a multiplication, instead of performing an algorithm to find its product. This CAT, with a worksheet format, is meant to assess whether students can identify the components of a multiplication and understand what is behind the multiplication algorithm.

Here you see a multiplication problem. You do not have to calculate the result. You only have to fill in the blanks.

**24 × 53** means that you have to calculate

- a)  $3 \times 4$  and  $3 \times 20$  and  $\_ \times 4$  and  $\_ \times 20$   
b)  $4 \times \_$  and  $4 \times \_$  and  $\_ \times 3$  and  $20 \times 50$   
c)  $20 \times 50$  and  $20 \times \_$  and  $4 \times 50$  and  $\_ \times 3$

Figure 2. CAT-2 breaking down a multiplication, a CAT with the format of a worksheet for assessing students' understanding in the domain of multiplication.

A teacher guide for using CATs and two two-hour meetings were arranged to support the teachers in implementing the CATs. The teachers filled in feedback forms after carrying out the CATs and wrote a final report about their use of the CATs. Their responses were used to determine whether they got insights from the CATs. When teachers referred to the mathematical content, a CAT aimed to assess and, either described specific information about their students or emphasized the novelty of the gained information, or referred to a fitting instructional adaptation, it was considered as evidence of gained insight. Moreover, 198 students' test scores on three district tests were analysed in order to explore the relationship between teachers' having gained insights from using CATs and the changes in their students' mathematics achievement.

In total, 193 responses about teachers' use of the CATs were collected, among which clear evidence of gained insight was identified in 57 responses (30%). Depending on the number of CATs from which a teacher gained insight, the teachers were divided into three groups. Five teachers formed the High Insight group, since evidence of gained insight was found in their responses about five or more CATs; fourteen teachers belonged to the Some Insight group, as evidence was found in three or fewer CATs; the remaining six teachers consisted of the No Insight group, because in their responses no evidence was found. The High Insight teachers could understand the idea of the CATs and use them to know more about their students' learning. Overall, these teachers favoured the

revelatory capacity of the CATs that differed in a specific way from the teaching tasks they would generally provide.

In contrast, the No Insight teachers often simply reported whether their students could solve the problems in general terms. Several of them considered CATs to be (too) similar to what is in the textbook and did not want to repeat what they had already taught, while others emphasized the CATs were too different, or too difficult, compared to their regular teaching. When looking at the students' scores on the three tests, no significant effect of having teachers' gained insights on student achievement was found.

## CONCLUSIONS

In this paper, we describe four studies that have been done in the project Improving Classroom Assessment in China (ICA-C). The most unexpected finding across different studies is that, although the involved Chinese primary school mathematics teachers generally endorsed the idea of conducting an assessment for improving teaching, adapting further teaching based on the assessment information seems to be a missing piece in their picture and practice of classroom assessment. Such practice was hardly mentioned in the teacher-written papers (*Study 1*) and was not found after the teachers got new information from using CATs (*Study 3*). In fact, many teachers reported assessment did not influence their teaching (*Study 2*). Classroom assessment can only function formatively when the collected information is actually used by teachers to adjust teaching to meet students' needs (Black & Wiliam, 1998). Otherwise, its promise to improve students' learning is impaired (Heritage, Kim, Vendlinski, & Herman, 2009). In this sense, the teachers in our studies did not really use assessment in a formative way, as their teaching was not shaped by the evidence of students' learning elicited by assessment. This may have to do with the fact that Chinese mathematics teachers pay more attention to their skills in providing information to students than gathering information about students (Cai & Wang, 2010).

Regarding the possibility for improving Chinese teachers' classroom assessment, it was found that CATs, as a domain-specific operationalisation that imply a strong approach to formative assessment, were feasible for teachers to carry out (*Study 3*). More importantly, teachers could gain new insights into their students' understanding from using CATs (*Study 4*). Thus, this research suggests it would be helpful to provide teachers with feasible techniques as a starting point to conduct a formative assessment. However, taking into account that the teachers often explained to their students how to solve the problems in the CATs before assessing them and the teachers did not use the elicited assessment information to make further instructional adaptation (*Study 3*), it seems that a fundamental shift in their conception of formative assessment is necessary before they can use assessment formatively and can make the most of their classroom assessment.

## ACKNOWLEDGEMENT

This research was funded by China Scholarship Council, grant number 201206860002; and the Netherlands Organization for Scientific Research, grant number NWO MaGW/PROO: Project 411-10-750. : The authors thank Prof. Lianhua Ning from Nanjing Normal University, China, for helping to contact schools, and all the teachers involved in this research project for their cooperation and contribution.

## REFERENCES

- Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction, 49*, 96–102.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.
- Brown, G. T. L., Hui, S. K., Yu, F. W., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5), 307–320.
- Cai, J., & Wang, T. (2010). Conceptions of effective mathematics teaching within a cultural context: Perspectives of teachers from China and the United States. *Journal of Mathematics Teacher Education, 13*, 265–287.
- 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.
- Curriculum Development Council (CDC). (2002). *Basic education curriculum guide Building on strengths (Primary 1–Secondary 3)*. Hong Kong: Author.
- De Lange, J. (1999). *Framework for classroom assessment in mathematics*. Madison, WI: NICLA/WCER.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24–31.
- Ministry of Education of the People's Republic of China (MoE) (2001). *Jichu jiaoyu kecheng gaige gangyao (shixing)* [Curriculum reform outline of basic education (trial version)]. Retrieved from <http://moe.edu.cn/publicfiles/business/htmlfiles/moe/s8001/201404/167343.html>.
- Ministry of Education of the People's Republic of China (MoE) (2011). *Yiwu jiaoyu shuxue kecheng biao zhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from [http://nct.moe.edu.cn/2014/curriculumPolicy\\_1115/3175.html](http://nct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html).
- National Council of Teachers of Mathematics (2013). *Formative Assessment: a position of the National Council of Teachers of Mathematics*. In NCTM (Ed.).
- Ni, Y. J., Li, Q., Li, X.Q., & Zou, J. (2011). China's new millennium curriculum reform in mathematics and its impact on classroom teaching and learning. In T. D. Huang &

- A. W. Wiseman (Eds.), *The impact and transformation of education policy in China* (pp.99–124). Bingley, UK: Emerald.
- Parkes, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed), *Sage handbook of research on classroom assessment* (pp.107–123). Thousand Oaks, CA: Sage.
- Ruthven, K. (1994). Better judgment: Rethinking assessment in mathematics education. *Educational Studies in Mathematics*, 27(4), 433–450.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2017, September 12–14) *Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment*. Paper presented at the NCME Special Conference on Classroom Assessment and Large-Scale Psychometrics: The Twain Shall Meet, Lawrence, KS.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014a). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE*, 9(1), e86817.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014b). 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 the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 5, pp.329–336). Vancouver, Canada: PME.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2019). *Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement*. Manuscript submitted for publication.
- 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.
- Zhao, D., Mulligan, J., & Mitchelmore, M. (2006). Case studies on mathematics assessment practices in Australian and Chinese primary schools. 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* (pp.261–276). New York, NY: Springer.
- 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.
- Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2018). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education*, 16, 1387–1407.
- Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2016). Teachers' use of classroom assessment techniques in primary mathematics education – An explorative study with six Chinese teachers. *International Journal of STEM Education*, 3(1), 19.
- Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2019). *Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques*. Manuscript submitted for publication.