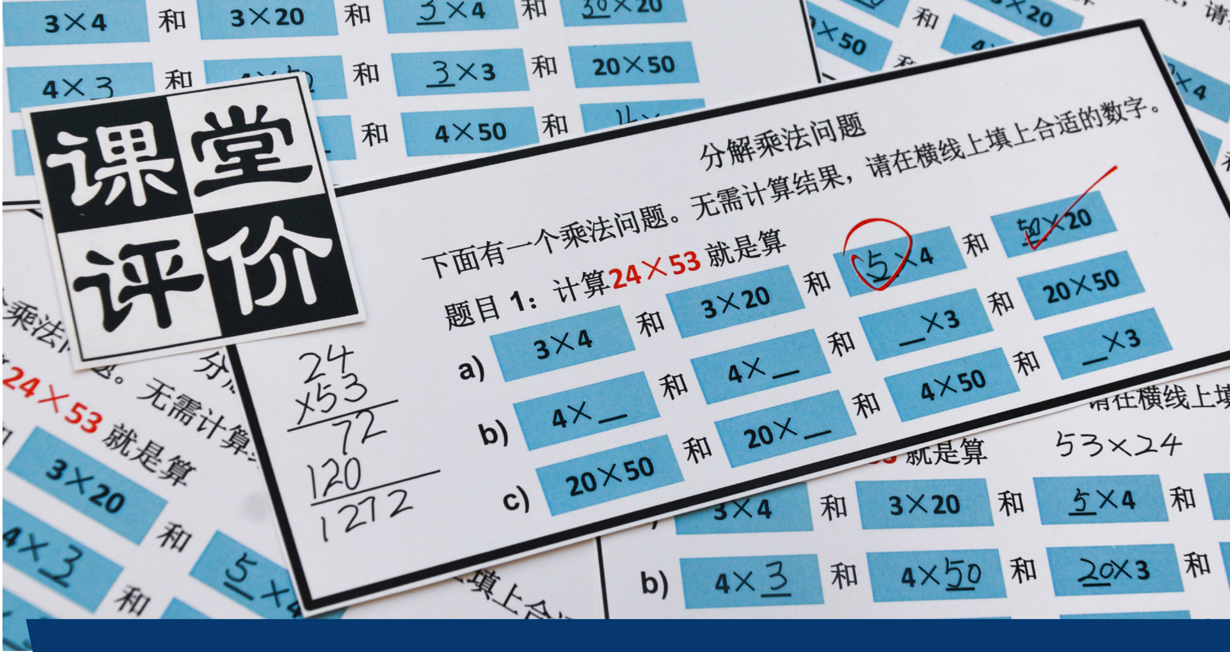


FI Scientific Library

Xiaoyan Zhao

Classroom assessment in Chinese primary school mathematics education

Faculteit Bètawetenschappen **FI**

Classroom assessment
in Chinese primary school mathematics education

The research in this thesis was supported by the China Scholarship Council under Grant 201206860002, and by the Netherlands Organization for Scientific Research in the Improving Classroom Assessment (ICA) project (NWO MaGW/PROO: Project 411-10-750).

This research was carried out in the context of the Dutch Interuniversity Centre for Educational Sciences.

ico

Zhao, Xiaoyan

Classroom assessment in Chinese primary school mathematics education / X. Zhao – Utrecht: Freudenthal Institute, Faculty of Science, Utrecht University / FI Scientific Library (formerly published as FIsme Scientific Library, CD-β Scientific Library) no. 96, 2018

Dissertation Utrecht University. With references. Met een samenvatting in het Nederlands. 附中文概要.

ISBN: 978-90-70786-40-3
Keywords: classroom assessment / classroom assessment techniques / mathematics education / primary education / China

Cover design: Vormgeving Faculteit Bètawetenschappen
Cover illustration: Xiaoyan Zhao
Printed by: Xerox, Utrecht

CLASSROOM ASSESSMENT
IN CHINESE PRIMARY SCHOOL MATHEMATICS EDUCATION

TOETSGEBRUIK BIJ REKENEN-WISKUNDE
IN HET CHINESE BASISONDERWIJS

(met een samenvatting in het Nederlands)

中国小学数学教育的课堂评价

(附中文概要)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 7 maart 2018 des middags te 12.45 uur

door
Xiaoyan Zhao

geboren op 1 november 1985
te Zhangjiakou, Hebei, China

Promotor: Prof. dr. M. H. A. M. van den Heuvel-Panhuizen
Copromotor: Dr. M. Veldhuis

For Jiakun (佳琨)

Table of contents

Chapter 1

Introduction

Introduction

1. Studying Chinese primary school mathematics teachers' use of assessment

In China, one of the most famous and widely spread teaching principles advocated and applied by Confucius (551 BC – 479 BC) is Yīn Cái Shī Jiào (因材施教), which means teachers should teach their students in accordance with their individual aptitudes, abilities, and situations. In this sense, the idea of using assessment (Píng Jià, 评价) to inform teaching has always been valued in the Chinese society. However, the wake-up call of putting assessment into the hands of teachers for the purpose of improving teaching and learning did not occur till the end of last century. Due to the long history of the civil service examination aimed at selecting officials in imperial China, for a long time assessment has mainly been associated with formal examinations and only the selective purpose of assessment is addressed (Li, 2000). This eventually led to the examination culture which resulted in all sorts of harmful effects on students' mathematics learning (Leung, 2008). To reduce this overemphasis on using assessment for the purpose of ranking and selection, in 2001, a new approach to assessment aimed at supporting teaching and learning was formally launched by the Ministry of Education (MoE, 2001a), in mainland China. According to the mathematics curriculum standards (MoE, 2001b, 2011), the main purpose of assessment is getting the whole picture of process and outcomes of students' mathematics learning, stimulating students to learn, and improving teachers' instruction. Teachers are suggested to employ various assessment methods to assess students' basic knowledge and skills, mathematical thinking, problem solving, and learning attitude. Moreover, teachers are recommended to report assessment results in a manner that can facilitate students to learn and to make most of assessment results by adapting instruction to cater for students' need.

This new approach to assessment is targeted at a population of, according to the national educational statistics data in 2015 (MoE, 2016), around 1.7 million Chinese primary school mathematics teachers to put into practice in 2.6 million classes. Even though the centralized educational system in mainland China has its advantage in delivering the important messages from top to bottom (Weng, 2013) and great endeavor has been made to help teachers to perceive and carry out this new approach to assessment since 2001 (Zhang, 2009), there is no guarantee that the improvement-oriented assessment initialized at the national level has been implemented faithfully at classroom level. The emerging question is whether, or to what extent, this idea has reached the classroom in primary schools.

When zooming in on teachers' use of assessment in primary school mathematics education, only a limited number of studies have been found. The overall finding is that, while they favor the idea of using assessment to improve teaching and learning (Brown, Hui, Yu, & Kennedy, 2011), mathematics teachers in primary schools need more help to enhance their assessment practice (Cai & Wang, 2010; Zhao, 2013; Zhao, Mulligan, & Mitchelmore, 2006). Furthermore, Zhong (2012) pointed out several problems in teachers' daily assessment practice. For example, teachers emphasize assessing what and how they teach but overlook assessing what and how students have learned. So they tend to assess students in a summative manner by focusing on whether students get the correct standard answer, but they give up the opportunity for formative use of assessment by avoiding students making mistakes. Nowadays, these problems still remain severe, although, since 2012, more and more teachers have been trying to improve their assessment practice (Zhong, 2017). To draw more attention to this issue, recently, the list of existing problems in assessment at classroom level addressed by Zhong was republished in *China Education Daily*, the newspaper sponsored by MoE.

Besides this message from the academic community, primary school mathematics teachers' voices and thoughts about their use of assessment can be found in papers published in professional teacher journals. Teachers' reflections and discussion are mainly about distinguishing various occasions in class, like when low achievers answer a question or when creative answers appear, and collecting good examples of what they can reply as teachers. Especially, how to stimulate learning by giving students appropriate praise is a hot topic. While many teachers share their experience of using this type of positive assessment to encourage students to learn, others express their doubt and confusion. For example, one teacher (Jiang, 2012) noticed that extensive misuse of praise takes place in class and many teachers seem to misunderstand the new approach to assessment.

The aforementioned description makes it clear that studies focusing on Chinese primary school mathematics teachers' use of assessment, in particular, how they perceive and practice in terms of assessment and how to support their assessment activities, are badly needed. When thinking about setting up such studies, two difficulties were encountered. First, only few relevant studies have been done in this area. The second difficulty has to do with my own experience from investigating teachers' assessment activity in secondary education during my master's study. By conducting a survey and analyzing two video-taped mathematics lessons, I found that although high school mathematics teachers considered assessment of students' learning to be an integral part of their teaching, in practice they were busy explaining the mathematics content and provided few opportunities for students to show what they already knew or understood. As a researcher, I could point out that teachers lacked initiative to elicit students' learning information, and that more attention should be paid to designing probe questions for assessing students. But I, in hindsight, was not able to provide any concrete examples of a good probe question, let alone make actionable suggestions about how to design them. Eventually, most of the discussion in my master's thesis went to how to provide feedback. In the face of such difficulties, it would be wise to take a worldwide perspective for studying Chinese primary school mathematics teachers' use of assessment and its improvement.

2. Improving Classroom Assessment (ICA) project in the Netherlands

The Netherlands is one of the countries in which a reform took place in mathematics education and at the same time the consequences regarding assessment were examined. This movement in Dutch mathematics education has been mainly guided by the theory of Realistic Mathematics Education (RME), which can be traced back to the 1960s (e.g., Freudenthal, 1991; Van den Heuvel-Panhuizen, 2000; Van den Heuvel-Panhuizen & Drijvers, 2014). In RME, the idea of mathematics as a human activity is addressed (Freudenthal, 1968). Students play an active role in learning; they can, by using contexts and models, pass through various levels of mathematization and develop their own mathematics (Van den Heuvel-Panhuizen, 1996). Teachers should provide students guided opportunity to re-invent mathematics by doing it (Van den Heuvel-Panhuizen, 2000). These views on mathematics, how students learn, and how teachers teach mathematics in RME led to developing a new approach to assessment in both secondary (De Lange, 1987) and primary education (Van den Heuvel-Panhuizen, 1996) in the Netherlands. Rather than ranking and selecting students, this new assessment in RME is part of teachers' daily teaching with the purpose of supporting teaching and learning, which is often called classroom assessment (De Lange, 1999) or didactical assessment (Van den Heuvel-Panhuizen, 1996).

When focusing on RME-based assessment in primary education, the main emphasis is on the problems used in the assessment (Van den Heuvel-Panhuizen, 1996). In the first place, it is crucial to determine what is assessed. Mathematical-didactical analysis (Treffers, 1980) plays an important role in this process to cast light on the significant points of learning particular mathematical content. Moreover, there are two requirements for problems to be suitable for RME-based assessment: they have to be meaningful and informative (Van den Heuvel-Panhuizen, 1996, p. 88). The former means the problems need to be accessible and worthwhile solving from students' perspective and reflect important learning goals in mathematics. The latter means the problems need to provide teachers with rich information about students' knowledge and skills for making decisions about the next teaching step. Around these two requirements, many more practical suggestions are, for example, using unfamiliar problems situations to assess students' high-order reasoning, using context in assessment problems, and the potential of open-ended problems.

With all the accumulated expertise and experience, in recent years, the Improving Classroom Assessment (ICA) project was set up, with the financial support from The Netherlands Organization for Scientific Research (NWO), to investigate primary school teachers' classroom assessment practice in mathematics, with a particular focus on possibilities for its improvement. Specifically, *classroom assessment* is the assessment that teachers continuously do during teaching to get access to students' skills and understanding in order to improve further instruction to meet students' learning needs (Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2014b, 2017). This is a formative approach to assessment (e.g., Black & Wiliam, 2009). For picturing Dutch teachers' classroom assessment practice in mathematics in primary education, a questionnaire was developed and a large-scale online survey was conducted nationwide. It was revealed teachers used various assessment methods for a wide range of purposes (Veldhuis, Van den Heuvel-Panhuizen, Vermeulen, & Eggen, 2013) and four assessment profiles of teachers could be identified: Enthusiastic assessors, Mainstream assessors, Non-enthusiastic assessors, and Alternative assessors (Veldhuis & Van den Heuvel-Panhuizen, 2014a).

After having acquired knowledge about what Dutch primary school teachers do in their assessment of students, improving teachers' assessment activities by the means of classroom assessment techniques was investigated. *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 (Veldhuis & Van den Heuvel-Panhuizen, 2014b, 2017). When designing the CATs, analyzing the textbooks teachers use was considered as the starting point, in order to make assessment close to teachers' 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 from a different perspective rather than what offered in the textbook. Finally, the format of CATs, for example a whole-class response system like red/green cards or a worksheet, was decided in order to help teachers gather students' information in an efficient and effective way. Following the abovementioned design principles, a series of CATs were developed to help Dutch teachers in Grade 3 to quickly find out information about students' understanding in the domain of number sense and number operations. To support the participating teachers to use CATs, a number of workshops were organized. Results from the studies show both teachers and students consider CATs to be feasible and useful (Veldhuis & Van den Heuvel-Panhuizen, 2014b). Also, positive effects of using the CATs, as a domain-specific operationalization of formative assessment, were found on Dutch students' mathematics achievement (Veldhuis & Van den Heuvel-Panhuizen, 2014b, 2017).

3. Improving Classroom Assessment in China (ICA-C) project

The Improving Classroom Assessment in China (ICA-C) Project was an extension of the ICA project in the Netherlands. In 2012, ICA-C was funded by the China Scholarship Council (CSC) with the aim of updating the knowledge about Chinese primary school mathematics teachers' perception and practice of classroom assessment, and exploring the possibility to help them improve their assessment activities by using CATs. In total, four studies were set up in the ICA-C project.

Since the meaning of, and the value people attach to assessment, are inevitably enmeshed in the wider culture of the community and differ between societies (Leung, 2008), what *classroom assessment* means and looks like for a primary teacher from the western world may be different from what Kè Táng Píng Jià (课堂评价) (literal translation of classroom assessment in Chinese) evokes in the mind of a Chinese primary school mathematics teacher. Particularly, the influence of the recent Chinese assessment reform in mathematics education has to be taken into account. Therefore, in the ICA-C project two studies were set up to shed light on how Chinese mathematics teachers in primary education consider and perform assessment: in the first study existing resources of teacher-written papers were examined and in the second study new data via an adjusted version of the questionnaire used in the ICA project were collected.

With the gained knowledge of the current situation of classroom assessment in mathematics in primary schools, we explored whether CATs could help Chinese teachers to enhance their assessment practice. The answer seems to be positive taking into consideration that, firstly, the conceptualization of CATs is quite in line with the new approach to assessment promoted in the Chinese assessment guidelines. Using CATs could provide Chinese teachers with concrete experiences of how to employ questioning to gauge students' mathematical understanding and to use students' learning information to improve teaching and learning. Moreover, the fact that features of CATs have close connections with the textbook may increase the chance for Chinese teachers to benefit from using them, since textbooks serve as their key resource to organize daily teaching (Cai, Ding, & Wang, 2014; Cai & Wang, 2010; Li, Chen, & Kulm, 2009). In addition, the formats of CATs can invite more students to engage in the assessment activities, which enables teachers to collect more information and get a better overview of students' learning in a big-class-size classroom. On top of that, positive findings of using CATs to improve teachers' assessment activities and students' achievement have been shown in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014b, 2017) and elsewhere (e.g., Andersson & Palm, 2017; Phelan, Choi, Niemi, Vendlinski, Baker, & Herman, 2012). However, it is also important to note that what would be a good approach to classroom assessment may differ in countries with different teaching approaches (e.g., Shepard, 2000), and clear differences in mathematics education can been recognized between East Asian countries and Western countries (Leung, Graf, & Lopez-Real, 2006). In short, it remains unsure whether CATs could be helpful in the context of Chinese primary school mathematics education. Therefore, two studies were set up in the ICA-C project to explore how Chinese primary school mathematics teachers use CATs and what new insights they could gain into their students' mathematical understanding from using CATs.

4. Structure of this PhD thesis

This PhD thesis is based on the four studies in the ICA-C project, and consists of a number of journal papers formatted as chapters. Table 1 illustrates the structure of the thesis, with title and purpose provided for each chapter.

Table 1
Structure of this PhD thesis

Chapter	Title	
1	Introduction	
2	Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers	Investigating Chinese primary school mathematics teachers' perceptions of classroom assessment
3	Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey	
4	Teachers' use of classroom assessment techniques in primary mathematics education—an explorative study with six Chinese teachers	Exploring whether CATs have the potential to help Chinese primary school mathematics teachers in their assessment practice
5	Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques	
6	Summary and discussion	

Chapter 2 reports on a review study in which primary school mathematics teachers' perceptions of classroom assessment were investigated. By examining 266 teacher-written papers addressing classroom assessment from the China National Knowledge Infrastructure database and published in 2011 and 2012, the following research questions are addressed:

- What do teacher-written papers reveal about Chinese primary school mathematics teachers' conceptions of classroom assessment? More specifically, for what purpose do teachers use assessment, what content is assessed, who acts as an assessor, which assessment methods are used and how are the assessment results reported and used?

- How are the teachers' conceptions of classroom assessment, as reflected in teacher-written papers, related to the assessment guidelines as included in the mathematics curriculum standards released in 2011?

Chapter 3 describes a large-scale questionnaire survey to identify Chinese primary school mathematics teachers' assessment profiles. The questionnaire used in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014a; Veldhuis, Van den Heuvel-Panhuizen, Vermeulen, & Eggen, 2013) was adapted to fit to the Chinese context for collecting information about teachers' backgrounds, their general teaching practice, their assessment practice, and their beliefs on assessment. Responses of 1101 teachers from 12 Chinese provinces and regions were analyzed to answer the following research question:

- What assessment profiles can be identified in Chinese primary school mathematics teachers?

Chapter 4 describes an explorative study, in which six female third-grade mathematics teachers from Nanjing, China, were provided with a series of CATs that were meant to assess their students' understanding of division. The following research questions are addressed:

- How are the CATs used in the context of Chinese primary school mathematics education?

- What information do the teachers, who use the CATs, get from CATs and what do they do with this information?

- Do these teachers think CATs are useful and do they want to use CATs in the future?

Chapter 5 reports on an intervention study in which special attention was paid to examine what new insights Chinese teachers gained into their students' mathematical understanding from using CATs. In this study, twenty-five teachers were offered eight CATs aimed at assessing students' understanding of multiplication. When teachers referred, in their responses about using CATs, to the mathematics 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, students' test scores on three district mathematics tests were collected in order to explore the relationship between teachers having gained insights from using CATs and the changes in their students' mathematics achievement. The research questions are as follows:

- What new insights can Chinese primary school mathematics teachers get about their students' understanding of mathematics by using CATs?

- Are insights into student learning the teachers gained from using CATs related to changes in their students' mathematics achievement?

Chapter 6 provides a summary of the findings of the four studies and discusses some practical implications. Moreover, suggestions for further research into classroom assessment are proposed.

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.

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.

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

Cai, J., Ding, M., & Wang, T. (2014). How do exemplary Chinese and U.S. mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*, 265–280.

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.

De Lange, J. (1987). *Mathematics, Insight and Meaning* (Doctoral dissertation). Utrecht: OW&OC, Utrecht University.

De Lange, J. (1999). *Framework for classroom assessment in mathematics*. Madison, WI: NICLA/WCER.

Freudenthal, H. (1968). Why to Teach Mathematics so as to Be Useful. *Educational Studies in Mathematics, 1*, 3–8.

Freudenthal, H. (1991). *Revisiting Mathematics Education. China Lectures.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Jiang, H. (2012) Ketang pingjia ruhe cong xujia zouxiang zhenshi [How can classroom assessment change from fake to real]. *Teacher, 18*, 41.

Li, Y. (2000). Purpose, function and types of mathematics assessment in China: A historical perspective. *ZDM Mathematics Education, 2000*, 192–200.

Li, Y., Chen, X., & Kulm, G. (2009). Mathematics teachers' practices and thinking in lesson plan development: a case of teaching fraction division. *ZDM Mathematics Education, 41*(6), 717–731

Leung, F. K. S. (2008). In the books there are golden houses: mathematics assessment in East Asia. *The International Journal on Mathematics Education, 40*(6), 983–992.

Leung, F. K. S., Graf, K. D., & Lopez-Real, F. J. (Eds.). (2006). *Mathematics education in different cultural traditions — a comparative study of East Asia and the West*. New York: Springer.

Ministry of Education of the People's Republic of China (MoE) (2001a). *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) (2001b). *Quanrizhi yiwu jiaoyu shuxue kecheng biaozhun (shiyangao)* [Mathematics curriculum standards of nine-year compulsory education (trial version)]. Beijing, China: Beijing Normal University Press.

Ministry of Education of the People's Republic of China (MoE) (2011a). *Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

Ministry of Education of the People's Republic of China (MoE) (2016). *2015 nian jiaoyu tongji shuju [Educational statistics data in 2015]*. Retrieved from http://www.moe.gov.cn/s78/A03/moe_560/jytjsj_2015/2015_qg/index_3.html

Phelan, J., Choi, K., Niemi, D. N., Vendlinski, T., 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.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Treffers, A. (1980). Cito item dito – over leerdoelgerichte toetsen meten [Cito idem dito – on the criterion-referenced testing of measuremenet]. *Wiskobas-Bulletin, 9* (6), 81–99.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education*. Utrecht, the Netherlands: CD-ß Press / Freudenthal Institute, Utrecht University.

Van den Heuvel-Panhuizen, M. (2000). *Mathematics education in the Netherlands: A guided tour. Freudenthal Institute Cd-rom for ICME9*. Utrecht: Utrecht University.

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.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 521-525). London: Springer.

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. (2017). *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.

Weng, W. (2013). Difang Jiaogai Xinlu [The new way of local educational reform]. *China Economic Report, 10*, 74–77.

Zhang, D. (2009). *Xiaoxue ketang jiaoxue xingchengxing pingjia wenti yu duice yanjiu* [The problems and solutions of formative assessment of classroom teaching of primary school] (Master's thesis). Available from China National Knowledge Infrastructure.

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, S. (2013) *Cujin xuexi de ketang pingjia yanjiu* [a study on classroom assessment for learning] (Doctoral dissertation). Available from China National Knowledge Infrastructure.

Zhong, Q. (2012). Ketang pingjia de tiaozhan [The challenges of classroom evaluation]. *Global Education, 1*(10), 10–16.

Zhong, Q. (2017, May 11). Tiaochu ketang pingjia de “Xinbagu” [Jump out of the “Eight New Limitations” of classroom assessment]. *China Education Daily*. Retrieved from http://paper.jyb.cn/zgjyb/html/2017-05/11/content_478291.htm?div=-1

Chapter 2

Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers

Xiaoyan Zhao[a], Marja van den Heuvel-Panhuizen[a,b], & Michiel Veldhuis[b]
[a]*Freudenthal Institute, Faculty of Science, Utrecht University, the Netherlands*
[b]*Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*

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.

Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers

Abstract

In this paper we report on a review of papers written by teachers aimed at knowing more about teachers' perceptions of the current situation of classroom assessment in primary mathematics education in China. The review is based on 266 papers included in the China National Knowledge Infrastructure database. We found that the teacher-authors reflected various aspects of their classroom assessment practice, including the purpose of assessment, the content of it, the person who is the assessor, the assessment methods that are used, and the feedback that is provided. Most attention was paid to feedback; it seems many teacher-authors considered classroom assessment to be equivalent to feedback. In general, the conceived classroom assessment practice as described in the papers echoed well nearly all aspects that are advocated in the Chinese mathematics curriculum standards. The only aspect that was scarcely discussed in the papers was the use of assessment results to adapt and improve instruction.

Keywords: Classroom assessment; mathematics education; primary school; teachers; review of teacher-written papers; China

1. Introduction

Classroom assessment, considered as assessment in the hands of teachers for the purpose of informing teaching and learning, has been recognized and promoted in mathematics education all over the world. This important role of classroom assessment is also reflected in the mathematics curriculum reform and the accompanying assessment reform in China, launched in 2001. After more than a decade of reform, however, it is still unclear how mathematics teachers perform assessment in their classrooms. Gaining more knowledge about this can be achieved in different ways; our approach in this study was conducting a review of papers in teacher journals written by Chinese teachers addressing classroom assessment in primary school mathematics education. By analyzing these teacher-written papers, we aimed at casting light on the activities teachers use in the assessment of their students, and whether the reported practice is related to the assessment guidelines in Chinese curriculum documents.

In the remainder of this introduction, we will elaborate successively on the role of classroom assessment in mathematics education in general, the content of the mathematics curriculum reform in China, the accompanying assessment guidelines, and finally we will formulate our research questions.

1.1 Classroom assessment in mathematics education

In the last decades, many countries have reformed their mathematics education towards a curriculum which no longer solely focuses on knowing facts and carrying out routine skills, but also on understanding and higher-order skills such as reasoning, modelling, and problem solving (see, e.g., NCTM, 1989, 2000). This reform in mathematics education has also changed the view on assessing students' learning and called for a new approach to assessment corresponding to and serving these changes in curricula (Leung, 2008; Romberg, Zarinnia, & Collis, 1990). A new approach to assessment is required to make it epistemologically consistent with the didactics of mathematics (Van den Heuvel-Panhuizen & Becker, 2003). Assessment should correspond to the curriculum that is taught and the learning theory that is adhered to (Shepard, 2000). This means that in addition to students' knowledge and skills also their ability to solve more complex problems should be assessed, that not only the correctness of students' solutions should be the focus of assessment but also the strategies employed by students, and, finally, that assessment is seen as an on-going process integrated within instruction (e.g., Berry, 2011; Romberg et al., 1990; Shepard, 2000; Suurtamm, Koch, & Arden, 2010; Van den Heuvel-Panhuizen, 1996). An important characteristic of this new approach to assessment is the awareness that assessment should not only be assessment of learning but also for learning, that is formative assessment, meaning that assessment should inform teachers' instructional decision making and students' learning (Assessment Reform Group, 1999; Black & Wiliam, 1998a, 1998b; Stiggins, 2002). Formative assessment in the hands of teachers (Van den Heuvel-Panhuizen & Becker, 2003, p. 698) that is interwoven with instruction and fully integrated in the teachers' daily teaching practice is often called classroom assessment (e.g., Black & Wiliam, 1998b; Brookhart, 2004; De Lange, 1999; Shepard, 2000; Stiggins & Chappuis, 2005; Wiliam, 2007). In general, classroom assessment includes all kinds of formative assessment in which the teacher has the lead. This means that the teacher makes decisions about when, for what purpose, and by which method information about students' learning is gathered with the aim of informing further steps in his/her teaching. To gather this information, teachers can use a variety of methods: ranging from observing students' problem solving, listening to students' answers to questions and quizzes, to examining students' written work and administering tests (Keeley & Tobey, 2011; Wiliam, 2011a). Also, as part of classroom assessment, teachers can offer students opportunities for carrying out self- and peer-assessment (e.g., Wiliam, 2011b), in which teachers need to carefully set up and manage the activity while students play the leading role in collecting and using assessment information for their own learning improvement (Andrade, 2010; Topping, 2010).

In accordance with the worldwide reform of mathematics education towards using assessment to enhance teaching and learning, in the last decades, the relevance of classroom assessment is increasingly acknowledged in many countries (Berry, 2011). In addition, the interest in classroom assessment was particularly triggered by the review done by Black and Wiliam (1998b) in which they revealed that teachers' use of classroom assessment can lead to the improvement of students' mathematics achievement. Although this review and its conclusions were criticized, particularly on the reported effect sizes (e.g., Bennett, 2011), enough evidence remains that teachers' use of classroom assessment is linked to an increase in students' learning (e.g., Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012; Kingston & Nash, 2011; McMillan, Venable, & Varier, 2013; Veldhuis & Van den Heuvel-Panhuizen, 2014). Therefore, policymakers have embraced the use of assessment for learning. For example, the US National Council of Teachers of Mathematics (NCTM, 2013) has strongly endorsed the integration of such assessment in daily instruction. In Hong Kong, the Curriculum Development Council (CDC, 2002) recommended that all schools should review their current assessment practices and put more emphasis on assessment for learning.

1.2 Mathematics curriculum reform in China

In mainland China, in 2001, the Ministry of Education of the People's Republic of China (MoE) initiated a curriculum reform with the purpose of better preparing students to meet the challenges of the 21st century by publishing a curriculum reform outline (MoE, 2001a). To help teachers, textbook designers, and other stakeholders in the nine-year compulsory education develop a clear view on the implementation of the curriculum reform in mathematics education, the MoE (2001b) also published in that same year the mathematics curriculum standards. In this document, one can find detailed information about (1) fundamental ideas about mathematics and mathematics education, and the structure of the mathematics curriculum standards, (2) the objectives of mathematics education in terms of knowledge and skills, mathematical thinking, problem solving, and mathematical and learning attitude, (3) mathematical content, and (4) suggestions with examples for instruction, assessment, and the design of mathematics textbooks and other materials.

The document of the mathematics curriculum standards (MoE, 2001b) was initially only used in parts of the country. By fall 2006 it became compulsory nationwide (Ni, Li, Li, & Zou, 2011). One year later, this was followed by the release of a revised version of the mathematics curriculum standards (MoE, 2007). This revised version was developed by a group of fourteen scholars, researchers, teacher educators, and expert teachers in mathematics education, organized and authorized by the MoE. The mathematics curriculum standards issued by MoE in 2001 were modified based on investigations into its use and the suggestions and critical remarks from mathematicians, experienced mathematics educators, and in-service mathematics teachers from more than ten provinces in China (Shi, Ma, & Liu, 2012). The latest version of the mathematics curriculum standards was published in December 2011 (MoE, 2011). In this version, it is emphasized that students should develop the ability of identifying and posing problems together with the ability of analyzing and solving problems. Moreover, it is stressed that attention should be paid to calculation, modelling, geometric visualization, and creativity, together with number sense, symbol sense, space concept, data analysis, reasoning, and application.

1.3 Assessment as described in the mathematics curriculum standards in China

China has a long history of examination-oriented education (Berry, 2011), which has been changed remarkably when in 2001, together with the curriculum reform, a new approach to assessment was promoted. In the curriculum reform outline (MoE, 2001a) it is mentioned that the assessment reform can be characterized by reducing the overemphasis on using assessment for differentiation and selection purposes, and using assessment to facilitate students' development, teachers' enhancement, and the improvement of the teaching and learning practice. The latest version of the mathematics curriculum standards (MoE, 2011) released some ten years later contains specific information about how assessment is conceptualized within the Chinese curriculum reform movement. To better support teachers' assessment practice, the mathematics curriculum standards document gives guidelines for the following aspects of assessment: (1) main purposes of assessment, (2) the content of assessment, (3) who can be an assessor, (4) the methods that can be used for assessment, and (5) suitable ways of reporting and using assessment results.

1.3.1 Purpose of assessment

In contrast to the use of assessment for differentiating and selecting students – which was common practice before the reform – the new approach to assessment is aimed at contributing to the teaching-learning process. In line with this, three purposes are mentioned in the mathematics curriculum standards: assessment should be used to get a comprehensive understanding of students' learning, to help students to enhance their learning, and to facilitate teachers to improve their instruction. However, the description of the purposes is very brief, and no further explanations or examples are given.

1.3.2 Content of assessment

For the content of assessment it is stipulated that it should address what mathematics students have to learn and what mathematical competences they have to develop. Table 1 shows the assessment guidelines and the two examples given for the competence domains in the mathematics curriculum standards of the nine-year compulsory education.

1.3.3 Assessor

Regarding the person who is doing the assessment, it is underlined in the mathematics curriculum standards that both the teacher, students themselves, their peers, and parents can participate as assessors. By establishing such a multi-actor system of assessment, both teachers' teaching and students' learning can be assessed. For example, by the end of a chapter the teacher can ask students to make a summary about their learning gains and difficulties. In this way the teacher can assess whether students have a good understanding of what is taught in the chapter. Additionally, with such a summary the students can reflect on what they have learned, the problems they have encountered, and how they could make improvements. Moreover, they can share and discuss their findings regarding their own learning and difficulties with their peers, which can make students learn from each other's experiences. If possible, parents are also welcome to join such assessment practice.

1.3.4 Method of assessment

Regarding the method of assessment, the document of the mathematics curriculum standards advises various methods including written tests, oral tests, open questions, activity reports, observations, interviews, exercises in and after class, and portfolios. Teachers also should understand the characteristics of all these different methods and choose an appropriate method that fits both the content to be assessed and their students' learning situations. The importance of written tests to assess students is explicitly emphasized, which also applies to the primary school grades. In connection with this, suggestions are given about how to design and conduct written tests.

Table 1

Assessment guidelines and examples for the competence domains included in the mathematics curriculum standards of the nine-year compulsory education in China (derived from MoE, 2011)

Competence domain	Assessment guideline	Example provided
Knowledge and skills	Assessment of mastery of basic knowledge and skills, which also includes assessing flexible application of these skills and students' understanding of mathematical concepts. For students who do not reach the objectives, providing extra instruction and more time is suggested.	By the end of Grade 3, students should be able to calculate within one minute 8 to 10 addition or subtraction problems up to 20, or 8 to 10 multiplication or division problems within the multiplication table.
Mathematical thinking and problem solving	Assessment of mathematical thinking and problem solving should be carried out by multiple methods during the whole process of mathematics learning, especially by using context problems.	In the second stage (Grade 4–6), teachers can provide the following task: Here is a 50 centimeter long string. Now use it to make a rectangle with sides of integers; and find when the rectangle has the biggest area. Teachers can provide questions with different difficulty levels to assess their students.
Mathematical and learning attitude	Assessment of students' mathematical and learning attitude should be carried out during the daily teaching, mainly using observations and activity reports. Also interviews can be carried out, to understand students' situation, their initiatives, interest, confidence, courage, and their cooperation and communication with others.	

1.3.5 Report and use of assessment results

The main guideline for reporting assessment results is that the feedback should contribute to the enhancement of students' confidence and their learning interests, help them to develop good learning habits and facilitate their learning. In the feedback to students, the assessment results should be provided descriptively combined with a grade or mark, with the focus on what the students learned, the progress they made, their potential, and where they need to improve. This feedback can be given orally or on paper. It is also emphasized that the process of getting feedback can be an emotional experience for students. Furthermore, the feedback should not only provide students with success experiences and boost their confidence, but should also let them know their weaknesses and where to improve.

About the use of assessment results it is only briefly described in the mathematics curriculum standards how teachers can benefit from the findings from an assessment. Based on the information about the students' learning level and their learning difficulties, it is suggested that teachers can adapt and improve their instruction. Finally, although the assessment promoted here is generally in alignment with the purposes and the characteristics of classroom assessment, the term classroom assessment is not employed in the mathematics curriculum standards.

1.4 Research questions

Over a decade ago the mathematics curriculum reform was launched in China. Whether and how the fundamental ideas of the reform were implemented in classrooms is an issue that lately has been receiving attention in the Chinese academic community. For example, Ni et al. (2011) investigated the impact of the curriculum reform on classroom teaching and learning, and found positive changes in teachers' beliefs, the cognitive level of the learning tasks provided to students, and the balanced development in students' mathematics achievement. Concerning the types of assessment teachers used, Ni et al. (2011) discovered that the teachers in the reform group were more able to employ reform-related types of assessment. However, other researchers have stated that assessment is a bottleneck for teachers (Zhu, 2012) and that teachers need to improve their classroom assessment (Zhong, 2012). Despite these findings about assessment, it is still largely unclear how mathematics teachers perform assessment in their classrooms, especially in primary school. To fill this knowledge gap, we set up the current study. We aimed to gain more knowledge, by examing teacher-written papers, on the activities teachers use in the assessment of their students, and whether this reported practice is related to the assessment guidelines in the curriculum documents. Even though there may be a profound gap between what teachers reported in their papers and what they factually did regarding to their assessment activity, what can at least be extracted from the teachers' publications are what assessment activities and views they have in their mind, which can be considered as a first requirement to use them in practice. Moreover, Herse (1979) stated that one's manner of presenting mathematics is an indication of what one believes to be most essential in it. Similarly, teachers' descriptions and views about what classroom assessment is and should be in their practices, can be a good resource to reveal what they believe to be of most importance in classroom assessment. Following Thompson (1984, 1992) who defines conception as a general mental structure encompassing beliefs, views, preferences, and the like, the research questions of our study were:

1. What do teacher-written papers reveal about Chinese primary school mathematics teachers' conceptions of classroom assessment? More specifically, for what purpose do teachers use assessment, what content is assessed, who acts as an assessor, which assessment methods are used and how are the assessment results reported and used?

2. How are the teachers' conceptions of classroom assessment, as reflected in teacher-written papers, related to the assessment guidelines as included in the mathematics curriculum standards released in 2011?

2. Method

2.1 Selection of teacher-written papers

To select papers we used the China National Knowledge Infrastructure (CNKI) database, which is the most comprehensive online resource for accessing China's intellectual output and includes journal papers as well as master and doctor theses, papers in proceedings, newspaper articles, and yearbooks. The selection of the papers (see Table 2) was carried out in July 2013.

Our goal was to collect teacher-written papers that cast light on the activities conducted by primary school mathematics teachers in the assessment of their students. Since a direct search for teacher-written papers is not possible in this database, we started with searching for papers in the Education category of CNKI related to *assessment*, in Chinese *PingJia* (评价), which is the literal translation of assessment. This first selection step resulted in 209,492 papers having *PingJia* in the title, keywords, or abstract. The earliest one was published in 1949, concerning evaluating the price of materials in military system (Jianli, 1949).

Because *PingJia* is a very broad term, which includes not only assessment at classroom level, but also all kinds of external evaluations – such as the evaluation of the quality of a school – a further selection was necessary to exclude forms of assessment that were not of interest in our review. In the second selection step a total of thirteen search words were used. These words were all related in some way to assessment conducted in the classroom such as classroom assessment, classroom teaching and learning assessment, assessment for learning, student assessment (including both self- and peer-assessment), learning assessment, and formative assessment. In this way we got 30,826 papers that address internal assessment, that is, assessment at the classroom level. To collect only the papers about mathematics education the third selection step was searching for papers with mathematics in the title, keywords, or abstract. The result was 2750 remaining papers. The fourth selection step was meant to only keep the papers about assessment in primary school. To optimize this search we did it in two steps: first the search option of the CNKI was used to search for primary school in the full papers and then the resulting papers were checked by a quick read. The remaining collection consisted of 904 papers that were downloaded from the CNKI database and put in EndNote.

As shown in Figure 1, the earliest paper referring to assessment at the classroom level in primary mathematics education was published in 1985. It is about assessing the quality of students' learning (Rong, 1985). In general, before 2001 very few papers were published about assessment at the classroom level in primary school mathematics education. In 2002, one year after the publication of the curriculum reform outline (MoE, 2001a) and the mathematics curriculum standards (MoE, 2001b), a notable increase in the number of published papers

Table 2

Steps in selecting teacher-written papers on classroom assessment in primary mathematics education in China

Papers in CNKI database belonging to the category Education		
Step	*Selection focus*	*Resulting papers*
1	Term 评价(assessment and/or evaluation) Search in title, abstract, and keywords	209,492 papers addressing: -all kinds of assessment
2	Terms 课堂评价 (classroom assessment), 课堂教学评价 (classroom teaching and learning assessment or classroom teaching assessment), 形成性评价 (formative assessment), 为了学的评价 (assessment for learning), 教师评价 (teacher assessment), 学生评价 (student assessment[a]), 作业评价 (assignment assessment), 即时评价 (immediate assessment), 延时评价 (delayed assessment), 过程评价 (assessment on process), 多元评价 (multiple assessment), 评价与提问 (assessment and questioning) and 评价与反馈 (assessment and feedback) Search in title, abstract, and keywords.	30,826 papers addressing: -internal assessment.
3	Term 数学 (mathematics) Search in title, abstract, and keywords.	2750 papers addressing: -internal assessment -mathematics
4	Term 小学 (primary school) Search in full paper	904 papers addressing: -internal assessment -mathematics -primary school

904 papers transferred to EndNote

Step	*Selection focus*	*Resulting papers*
5	Publication year 2011 or 2012 Search in the EndNote column Publication year	360 papers published in 2011 or 2012 addressing: -internal assessment -mathematics -primary school
6	Paper contains formative assessment conducted by the teacher, that is, contains classroom assessment Search in full paper	283 papers published in 2011 or 2012 addressing: -mathematics -primary school -classroom assessment
7	Paper is written by a teacher Search in author's affiliation	266 teacher-written papers published in 2011 or 2012 addressing: -mathematics -primary school -classroom assessment

[a] Student assessment includes self- and peer-assessment.

on assessment can be seen. A further boost took place in 2007 when a revised version of the mathematics curriculum standards was published (MoE, 2007). Since then the number of published papers showed a steady increase.

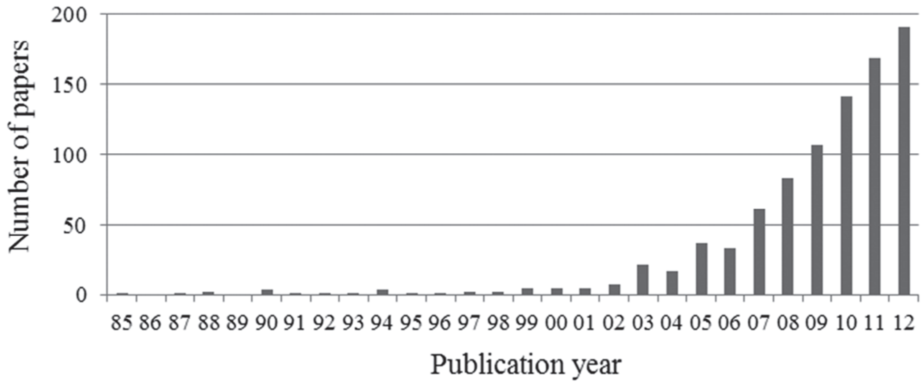


Figure 1. Frequency distribution of papers on assessment at the classroom level in primary mathematics education found in the CNKI database during 1985–2012

Since these papers still involved all kinds of assessment taking place in classrooms, a further selection was necessary. However, just searching for classroom assessment would not have given the intended selection of papers. As in the rest of the world, in China there are many interpretations of the term classroom assessment. According to Zhong (2012), classroom assessment covers both assessment of teachers' teaching and students' learning, whereas Wang (2011) considers classroom assessment as only appraising or correcting students' performance. Classroom teaching and learning assessment, one of the search words used in the second selection step, is also defined differently between authors, ranging from assessing teachers, to assessing students, to assessing the process and effectiveness of teaching, and to assessing teaching and learning as a whole (Cao, Li, & Qing, 2011). So, to guarantee that the papers in the review were about activities conducted by teachers to assess their students' learning, the papers had to be read more closely. However, doing this for 904 papers was not feasible. Therefore, we decided to include only the papers published in the two latest full years, 2011 and 2012, in our collection. This means that the papers in the review appeared one decade after the mathematics curriculum reform was launched in 2001. This fifth selection step brought us to 360 papers. Then, in the sixth selection step, the full papers were read to identify those that discuss classroom assessment, that is, formative assessment conducted by the teacher. This resulted in 283 papers. Finally, a seventh selection step was carried out in which the information about the authors' affiliations was used to identify the papers that were written by teachers. In fact, except for thirteen papers from researchers and four papers from consultants, the majority papers, 266 out of 283, were written by teachers, which were used for the review. A numbered list of the references of these papers is in Appendix A.

2.2 Coding framework

The development of a coding framework (see Table 3) for carrying out the review in a systematic way started with reading the first fifty papers published in 2011 to find out whether the aspects of assessment for which guidelines are provided in the mathematics curriculum standards could function as suitable categories to be included in the coding framework. It turned out that most of the aspects were indeed discussed in the papers, including the purpose of assessment, the content of assessment, the person who is the assessor, and the method of assessment. With respect to reporting and using of the assessment results, it was found that most of the papers addressed mainly the aspect of reporting, that is, the provision of feedback; whether and how the assessment results were used for adapting and improving teaching was rarely described in the papers.

After the main categories were established, the subcategories were determined. The suggested assessment guidelines in the mathematics curriculum standards were the first source for deciding possible subcategories. The other source was what was discussed by the teacher-authors. For example, according to the mathematics curriculum standards, three main purposes of assessment can be distinguished: checking students' understanding, stimulating students to learn, and informing teachers' instructional decision-making. Moreover, in the teacher-written papers, some teachers also talked about using assessment to create a harmonious classroom environment or to promote their students' confidence. So, these two were added as subcategories of the purpose of assessment in our framework.

Table 3
Coding framework for teacher-authors' conceptions of classroom assessment

Category	Subcategory	
Purpose	Checking students' understanding	
	Stimulating students to learn	
	Informing teachers' instructional decision-making	
	Establishing a harmonious classroom environment	
	Promoting students' confidence	
Content	Basic knowledge and skills	
	Mathematical thinking and problem solving	
	Mathematical and learning attitude	
Assessor	Teachers	
	Students (including assessing themselves and peers)	
	Parents	
Method	Observation	
	Questioning	
	Classroom discussion	
	In-class assignment	
	After-class assignment	
	Presentation	
	Portfolio	
	Quiz/written test	
	Mathematics diary	
Feedback	Focus	Task-related
		Process-related
		Person-related
	Nature	Only positive
		Balanced (positive and what needs to be improved)
	Mode	Verbal
		Written
		Body language
		Material incentives
	Timing	Immediate (during class)
		Delayed (after class)
Teaching adaptation		

Table 3 shows the subcategories that were used for the purpose of assessment, the content of assessment, the person who is the assessor, the methods of assessment, and the feedback given to the students. For the category of teaching adaptation no subcategory was provided, since it was only briefly described in the mathematics curriculum standards and was rarely discussed in the teacher-written papers. Examples for each subcategory and the category of teaching adaptation can be found in Appendix B. Almost all the categories and the subcategories mentioned in Table 3 are not exclusive. This means that multiple coding of the papers was possible. The subcategories were exclusive only with respect to the nature of the feedback (the feedback was either positive or balanced).

In order to code a paper with a *yes* for a particular category or a subcategory, two requirements from the coding protocol had to be met at the same time. The first requirement was that there must be a clear message from the teacher-author in which it is expressed, for example, in what way classroom assessment *should be* conducted. Such a message could be conveyed by the teachers' own general statements, by quotations from the mathematics curriculum standards, or by concrete examples from the classroom. However, negative statements or unfavorable examples alone were not coded, since saying what was not appropriate did not directly indicate what was advocated by the teacher. The other requirement was that a category or a subcategory should really be addressed by the author, rather than casually mentioned. Yet, the amount of attention that was required to consider a category or a subcategory as being addressed differed between categories. For example, how to use a mathematics diary to assess students could be treated in the whole paper, while the purpose of classroom assessment could be dealt with by only referring to it once at the beginning of the paper. Because feedback was mentioned in almost every paper and many of them provided no further explanation, coding these papers as dealing with feedback would not be informative. Therefore, the feedback in a paper was only coded if at least a quarter of the paper discussed feedback.

The coding was done by the first author of this paper. After the coding was finished, 10% of the papers (of each publication year, the first thirteen papers in the list of the teacher-written papers in Appendix A) were reviewed by a second coder not involved in this study. The interrater reliability was found to be substantial, with $\kappa = 0.73$. In the 26 papers, 86% of the codes assigned by the external coder were identical to the codes of the first author.

3. Results

3.1 Characteristics of the papers included in the review

Of the 266 teacher-written papers about classroom assessment in primary mathematics education that were analyzed, 123 were published in 2011 and 143 were published in 2012. The authors were from 26 out of the 31 provinces, municipalities, and autonomous regions in mainland China. More than half of the papers (58%) were written by teachers from the provinces Jiangsu, Zhejiang, and Fujian, which are the three provinces with the highest overall level of educational development (Wang, Yuan, Tian, & Zhang, 2013). Surprisingly no papers were found from authors from Beijing or Tianjin, which are the municipalities with the second and fourth highest overall level of educational development in mainland China (Wang et al., 2013).

The teacher-written papers mainly involve teachers' reflections on their own practice of classroom assessment or tips learned from others. The length of the papers ranges from about half a page to four pages and most papers have one or two pages, containing about 2000 to 3000 Chinese characters. These papers usually start with a statement on the importance of classroom assessment, a quotation from the mathematics curriculum standards, or an example of unfavorable assessment practice the teacher-authors encountered. Then, in the main part of the papers the teacher-authors mostly gave their own suggestions for conducting classroom assessment, illustrated by concrete examples. For example, feedback should pay attention to the strategies students used, and it is better to correct low-achievers' work face to face. In the conclusion part of the papers, the teacher-authors generally provided a short conclusion to highlight the main idea of the papers. In addition to quoting from the mathematics curriculum standards, many teachers quoted educators, psychologists, or celebrities both national – such as Xingzhi Tao – and international – such as Howard Earl Gardner or Abraham Lincoln – to support their ideas. The teacher-authors also used a variety of metaphors to describe classroom assessment, like the GPS to guide students' learning in class, the catalyst of teaching and learning, and the booster to facilitate learning. In 60 papers we found that classroom assessment was associated with beauty, art, or positive emotional feelings. Specifically, in 7 papers the teacher-authors talked about the beauty of the classroom assessment, like in #10 (the paper numbered 10 in Appendix A) it was said that "assessment is the most gorgeous flower in classroom teaching", and in #101 the teacher-author said that doing classroom assessment "turned out to be a beautiful story." In 14 papers it was mentioned that doing classroom assessment is a kind of art. Finally, in 39 other papers adjectives were used to describe emotion-related effect of classroom assessment, such as that classroom assessment makes the mathematics teaching or the class beautiful, glorious, energetic, poetic, charming, warm or sweet.

3.2 Teacher-authors' conceptions of classroom assessment

Table 4 shows that of the 266 teacher-written papers about classroom assessment in primary mathematics education, 131 papers (49%) mention the purpose of classroom assessment, 187 papers (70%) address the content to be assessed, all 266 papers (100%) discuss who the assessor is, 208 papers (78%) describe the used method(s), and in 198 papers (74%) at least a quarter of the paper refers to feedback. However, in only 9 papers (3%) it was found that the teacher-authors explicated making teaching adaptation based on the assessment information.

3.2.1 Classroom assessment purpose

In total, we found 131 papers referring to one or more purposes of classroom assessment. Stimulating students to learn is mentioned most often (in 81% of the papers). This purpose is followed by checking students' understanding (69%) and informing teachers' instructional decision-making (60%).

In addition, a few papers refer to using classroom assessment to establish a harmonious classroom environment and to promote students' confidence. These purposes of doing assessment are both mentioned in 3% of the papers. For example, in paper #9 (see Appendix A) the teacher-author wrote that she was "using immediate assessment to create positive emotion [in the class] and to create a harmonious and democratic classroom learning environment." Similarly, in #10 the teacher-author expressed it in the following way: "The purpose is to find students' strengths during their learning process, to give encouragement, and to create a harmonious environment for teaching and learning." In #192 the teacher-author was very clear that the purpose of assessment was to contribute to students' confidence. He even titled his paper as "Three approaches to using assessment to build students' confidence." According to him, "teachers are required to [...] help students to recognize themselves and build confidence."

Table 4

Number of papers referring to the different aspects and its subcategories of classroom assessment (N = 266)

Category	Subcategory	Number of papers	
Purpose		131	
	Stimulating students to learn		106 (81%)
	Checking students' understanding		90 (69%)
	Informing teachers' instructional decision-making		78 (60%)
	Establishing a harmonious classroom environment		4 (3%)
	Promoting students' confidence		4 (3%)
Content		187	
	Basic knowledge and skills		129 (69%)
	Mathematical thinking and problem solving		105 (56%)
	Mathematical and learning attitude		106 (57%)
Assessor		266	
	Teachers		265 (100%)
	Students (including assessing themselves and peers)		139 (52%)
	Parents		30 (11%)
Method		208	
	Exercises in class		129 (62%)
	Homework		93 (45%)
	Questioning		83 (40%)
	Observation		29 (14%)

Category	Subcategory		Number of papers		
Method	Discussion				26 (13%)
	Quiz/written test				25 (12%)
	Portfolio				23 (11%)
	Presentation				11 (5%)
	Mathematics diary				11 (5%)
Feedback			198		
	Focus			146	
		Task-related			115 (58%)
		Process-related			80 (40%)
		Person-related			72 (36%)
	Nature			198	
		Balanced (positive and what needs to be improved)			134 (68%)
		Only positive			64 (32%)
	Mode			196	
		Verbal			153 (77%)
		Written			76 (38%)
		Body language			44 (22%)
		Material incentives			26 (13%)
	Timing			194	
		Immediate (during class)			147 (74%)
		Delayed (after class)			79 (40%)
Teaching adaptation			9		

3.2.2 Content of classroom assessment

In 187 papers the aspect of what has to be assessed is addressed. Table 4 shows that most attention is paid to assessing students' basic knowledge and skills (69%), followed by assessing mathematical thinking and problem solving (56%) and assessing students' mathematical and learning attitude (57%). In addition to the different content that is assessed, in 82 papers (44%) the teacher-authors mentioned explicitly that the various aspects of students' competences are interrelated and therefore these competences should not be assessed in isolation but together. For example, the teacher-author of #195 stated that classroom assessment "should not only focus on the understanding and application of knowledge and skills, both also on emotion and attitude to facilitate students' overall, harmonious and sustainable development."

3.2.2.1 Basic knowledge and skills. We found that many teacher- authors wrote about the necessity and importance of assessing students' basic knowledge and skills, but only a few of them described explicitly how to assess this content; however, where papers contain more detailed information about how to assess basic knowledge and skills, the given examples mainly consist of bare number problems. Another finding was that some teacher-authors reflected a broad interpretation of basic knowledge and skills, meaning that they did not only focus on the correctness of the problems, but also tried to get information about how students solved them and whether they really understood them. For example, the teacher-author of paper #16 underlined that "the principle standard [to assess] is not to see whether they learned by heart a formula or can use the formula to calculate the right answers, but to see whether they know how the formula is developed" and the author of #98 "asked her students to describe how they solved the problem of 45-9".

3.2.2.2 Mathematical thinking and problem solving. The importance of assessing mathematical thinking and problem solving is recognized by many teachers, like the teacher-author of #44, who said: "The ultimate goal of students' learning is application of knowledge and skills they learned in real contexts. Therefore, after students have learned some new knowledge, it should be assessed not only whether they command the knowledge, but also whether they can use the knowledge to solve problems." Also, we found suggestions for how to assess this content. According to the teacher-author of #16, the teachers should design problems within a real life situation, like how much to pay when buying one hundred notebooks with discount. Providing open-ended problems was used in #98, in which the teacher asked her students to design a rectangular garden with a particular area. Furthermore, in paper #70, the teacher required his students to think of their own questions based on a frequency table and then solve their own problems posed on these data.

3.2.2.3 Mathematical and learning attitude. We found that many teacher-authors noticed that teachers need to keep an eye on students' learning interest, their initiatives and engagement in learning activities, and their communication and cooperation with others. To assess students' attitude several indicators were mentioned, such as how many students raised their hands to answer the teachers' questions, whether they showed selfconfidence when answering questions and whether they wrote their homework in a clean and neat way. The teacher-authors of #70 and #205 also paid attention to whether their students behaved well, like whether they started to interrupt while others were still talking. And the teacher-author of #137 mentioned that he/she assessed whether the students were sitting upright in class.

3.2.3 Assessor in classroom assessment

In all 266 papers except for one, it was found that the teacher is the person who conducted the classroom assessment. Nevertheless, in 139 papers (52%) students also play a role as assessor. Taking paper #6 as an example, the teacher-authors described that they "organized the students to correct their exercises together in a group to develop their sense of doing assessment." In #5, the teacher-author emphasized that "by using self- and peer-assessment, the function of classroom assessment will be largely magnified. The students are not just the passive assessees anymore; they take part in assessment actively and experience the joy of doing assessment."

Also, in 30 papers (11%), it was found that parents are encouraged to be involved as assessors. The teacher-author of #81 wrote: "[W]ithout the participation of the parents, the assessment of students is not complete. By asking parents to observe how their children learn mathematics at home and sending the assessment results back to the teachers, the teachers can get a better understanding of the students' learning and adjust their instruction instantly."

The teacher-author of paper #145 wrote that an assessment form had been designed for the parents to assess their children's performance of doing homework.

3.2.4 Method of classroom assessment

Table 4 shows that there are 208 papers discussing the method of classroom assessment in general and the methods that were found in these papers. The most frequently discussed method is doing exercises in class (62%). This method is followed by giving homework (45%) and questioning (40%). Less attention is paid to carrying out observations (14%), organizing a discussion (13%), doing quizzes (12%) and keeping a portfolio (11%). Rarely applied methods are asking students to do a presentation (5%) and write a mathematics diary (5%).

3.2.4.1 Exercises in class. Giving exercises as an assessment method can be done at the beginning of a lesson to check whether the students are ready for the new learning material. The teacher-author of paper #11 described that she "prepared ten problems of mental calculation to ask her students to solve before teaching calculation in a smart way." Also, exercises were used after the instruction of a certain topic. For example, the teacher-author of #60 used exercises in class "to know immediately to what extent the students have mastered the most important and most difficult part of the topic in this lesson." Similarly, in paper #160, after the students learned the decimals, the teacher gave the exercise: "Which '0' can be left out [without changing] the following numbers? (3.09, 0.300, 1.800, 500, 5.780 and 0.040)." When using exercises in class to assess students' learning a few teachers also provided open-ended exercises. For example, the teacher-author in #88 asked students to think of their own questions related to percentages and solve them given the situation that "there are 20 apples and 25 pears".

3.2.4.2 Homework. Written homework gives teachers several opportunities to get information about their students' learning. For example, in paper #127, the teacher-author described that based on the homework he found "one student made the same mistake when solving a series of problems: 46×54+46×46 = 46×3+54 = 192, 25×99+25 = 25×2+99 = 149, and 99×99+99 = 99×3 = 297." In #77, it was emphasized that "what students' homework looks like also could reflect students' learning habit and attitude." Furthermore, instead of assigning simple written homework, the teachers tended to provide various types of homework. Like the teacher-author of #3 asked the students to go to a supermarket and collect the prices of goods, and the teacher-author of #147 asked her students to estimate the distance between school and their homes.

3.2.4.3 Questioning. By asking students to answer specific questions, teachers can quickly find out where students are. Questions were usually posed to be answered by any student in the class, but only one or a few students were selected to give their answers. For example, in paper #66, the teacher provided a true-or-false question: "Cutting a round piece of paper into two parts, then one of the two parts will be 1/2 [of the piece of the paper], true or false?" She found that some students said "true" and some said "false". Also, a few teachers raised questions to be answered by the whole class at the same time. The teacher-authors of #214 described an example that "[s]tudents with a prime [student number] are required to raise their right hands; while those with a [student] number that has a divisor other than 1 and itself raise their left hands." In addition, teachers also mentioned that only asking questions that can be answered by a simple "yes" or "no" is not enough. For example, the author of paper #69 reported that teachers should use more "why" questions to look closer at students' thinking. Furthermore, the teachers noticed that the questions should be clear to the students, and sufficient time should be given to guarantee students can think carefully.

3.2.4.4 Observation. In their writings the teachers made it clear that the observation took place when students were involved in various activities, such as doing exercises, answering a question, and discussing with their neighbors. In these observations an eye could be kept on the students' mathematics learning by focusing on how many students solved a problem, what strategies were used, and what mistakes were made. The author of #83 emphasized that such observations require that teachers "observe their students through the lens of mathematics for finding information about students' mathematics learning." In addition, the observation was used to throw light on the students' attitude and social behavior. For example, by observing how a student raised her hand, the author of #64 noticed that the student was lacking in confidence to give her opinion. Furthermore, in a few papers, for example in #10 and #116, the teacher-authors made observation forms to guide their classroom assessment practice.

3.2.4.5 Discussion. The method of organizing a discussion was chosen because in this method, according to the teacher-authors, more students could be involved and engaged in the assessment. Moreover, discussions are considered to have the potential to reveal students' deeper thinking. In the papers we found both discussion in small groups and in the whole class. The topics of discussion mainly involve the key concept of a lesson, different strategies of solving a problem and students' mistakes. For example, the author of paper #8, after the instruction, assessed his students by letting them discuss the relationship between rectangle and square in groups. In paper #38, it is described how an experienced teacher organized an assessment by asking his students to discuss three different answers to the problem of factorizing 36.

3.2.4.6 Quiz/written test. From the papers it can be derived that quizzes are normally held after a period of learning, for example at the end of a chapter. Quizzes mostly have a broad coverage of assessed content and generally take about one lesson. Quizzes, contrast with the method of giving exercises which is often used immediately after the instruction, is mostly focused only on the topic taught in that lesson, and needs only a few minutes. In the papers, the teacher-authors agreed that doing quizzes is a powerful way to elicit information about students' mastery of basic knowledge and skills. However, they can also be used to assess students' deep understanding and ability to solve problems. For example, the authors of #201 and #232 suggested to reduce problems which demand rote memory and include well-designed context problems or open questions, while papers #16 and #61 recommended to provide quizzes with different difficulty levels and offer the students the opportunity to choose a quiz that fits their own situation.

3.2.4.7 Portfolio. The references made in the papers to keeping a portfolio as a method to assess students' learning make it clear that both teachers and students can decide what to put in the portfolios and that a wide variety of files can be collected, for example, a student's best homework, a creative solution to a problem, a summary of what a student learned and some typical mistakes a student made. In the papers we found two main reasons for using portfolios to assess students. According to the teacher-author of paper #132, using portfolios could help teachers to "track individual students' longitudinal development." Furthermore, the teacher-authors of paper #165 valued the advantage of using portfolios to improve students' self-assessment, because "when students are responsible for making decisions about what to put into their portfolio, they get the opportunity to reflect on their learning and assess themselves."

3.2.4.8 Presentation. Only a few papers mention asking the students to give a presentation as a way of assessing them. Yet teacher-authors who referred to this method emphasized that one can get deep insight into students' understanding in this way. In paper #213, the students were asked to do a self-study about a new topic and give a short presentation about what they have learned. This let the teacher find out whether this new topic was easy or difficult for the students. Another example is from the teacher-author in #220. He asked his students to give a presentation about their self-designed problems and their solutions.

3.2.4.9 Mathematics diary. The few teacher-authors who mentioned the mathematics diary are quite positive about this assessment method. Asking students to keep a mathematics diary is considered as a rich resource to reflect all aspects of students' mathematics learning. In their mathematics diaries, students can ask questions or explain their confusion, report mistakes they made and their corrections, describe phenomena they found in daily life which are related to mathematics, or tell a story about mathematicians. The teacher-author of #252 included part of a student's mathematics diary in the paper:

> After we learned calculation in a smart way, Teacher Chen reminded us to summarize where it is easily to make a mistake. I noticed I need to be more careful in two situations [namely, when removing parentheses or adding parentheses without changing the original problem]. At that moment, I thought how it was possible for me to make such simple mistakes. However, it turned out I did make an error in my homework. [...] It was just because I was careless. How ashamed I am!

The possibility to have a wide scope in this assessment method is reflected in paper #68. Here, the teacher-author made it clear that "based on students' mathematics diaries, teachers can assess students' learning about knowledge and skills. More importantly, teachers are able to discern students' learning interest and attitude as well."

In addition, in 36 papers we found that the teacher-authors mentioned designing tasks which included exercises in class and homework, mainly for students to practice more and to learn from these tasks. Only in 17 of these papers did the teacher-authors mention the design of questions and quiz items to assess their students. For both purposes they emphasized that more context problems and open-ended problems should be offered to students. Moreover, some teacher-authors pointed out that it is important to offer fun problems to students and to offer different problems to individual students.

3.2.5 Feedback

The aspect of assessment that is mostly addressed in the reviewed collection of teacher-written papers on classroom assessment is providing feedback to students. Although we only counted a paper as referring to feedback when substantial attention is paid to this aspect of assessment, this still was the case in 198 (74%) of the papers. Actually, in 89 of them nearly the whole paper is dedicated to the issue of giving students feedback.

A further finding was that in 64 papers the teacher-authors seemed to consider classroom assessment as equivalent to feedback. For example, according to the teacher-author of #144, assessment is “providing students with feedback about their learning situation.” The author of #158 made it clear that: “classroom assessment is, during the process of teaching and learning, the positive or negative judgements made by assessors based on assessees’ performance.” In paper #111 the teacher-author used interchangeably the terms 课堂评价 (classroom assessment) and 课堂评价语言 (classroom assessment language). She gave as an example of her assessment: “Cao is very shy in the discussion and she did not dare to express her thinking.” Also we found that when the teacher-authors reflected on their practice of classroom assessment, they sometimes did not give any information about the concrete questions they asked their students, but only mentioned their reactions after the students gave their answer. Like the author of #166 who summarized: “[W]hen students reply with a wrong answer, I can say ‘you are not far from the right answer’ or ‘I know you have thought it over, shall we first listen to other students?’ ” The same was done by the author of #119. When she described her assessment practice, she just summed up the possible reactions she could give, such as: “[T]o students who are careless, I can say ‘I hope you can say goodbye to carelessness and make friend with carefulness' or 'your handwriting is very beautiful. It is much better if you can be careful [while solving problems in homework]'."

With respect to the different aspects of feedback found in the papers, Table 4 shows that in the 146 papers that discuss the focus of feedback, the feedback is mostly task-related (58%), followed by process-related (40%) and person-related (36%) feedback. As for the nature, in the 198 papers that refer to this aspect of feedback we found fewer papers in which the teacher-authors only thought of giving positive feedback (32%), that is praising students, and more papers in which the teacher-authors reflected a balanced way of providing feedback (68%). The latter means that the feedback is not completely positive, but also contains clear information about what needs to be improved.

The mode of feedback was found to be addressed in 196 papers with verbal feedback (77%) as the dominant mode, followed by written feedback (38%) and feedback given by body language (22%). Using material incentives as feedback (13%) is mentioned only in some papers. In the 194 papers that give information about the timing of providing feedback we found that much more feedback was given during class (74%) than after class (40%).

3.2.6 Teaching adaptation

Only in nine papers, the teacher-authors mentioned how results of classroom assessment were used for adapting and improving their instruction. From the examples described in #38 and #161, we found that the teachers gave supplementary exercises immediately after finding their students could not answer the questions correctly. Teacher-authors of #86 and #219 described that teachers are required to analyze students' results on written tests and students' mistakes in order to adapt their further teaching. Similarly, in paper #234, the teacher-author recommended to analyze students' mistakes in their homework to make instruction meet students' needs. In the remaining papers, the teacher-authors mentioned that they thought of using or had used assessment results for instructional decision making.

3.3 Relation between assessment conceptions in the papers and the assessment guidelines

The papers revealed that the teacher-authors took the assessment guidelines in the mathematics curriculum standards as a source for their classroom assessment practice. Evidence for this could be found in 142 papers. The teacher-authors explicitly stated that their classroom assessment activities are in line with what is advocated in the curriculum reform. For example, the teacher-author of #94 wrote: “[T]he idea of assessment in the [2001] curriculum reform is student-development-oriented; I explored how to assess my students [based on this new idea] in my classroom teaching.” Other teacher-authors clearly paraphrased the guidelines or even exactly cited them. In total, in 118 papers literal quotations were found.

3.3.1 Purpose of assessment

Regarding the purpose of assessment, the document of the mathematics curriculum standards emphasizes that the main purpose of assessment is to (1) “get the whole picture of process and outcomes of students’ mathematics learning”, (2) “stimulate students to learn”, and (3) “improve teachers’ instruction” (see MoE, 2011, p. 33). All three purposes are also mentioned by the teacher-authors, with stimulating students to learn discussed most often. Furthermore, we found in 41 papers that the teacher-authors used the entire and exact wording of the assessment guidelines. In addition, the teachers also mentioned two other purposes, namely establishing a harmonious classroom environment and promoting students’ confidence. However, these were only found in a very few papers.

3.3.2 Content of assessment

The descriptions in the papers about the content that is assessed reflect that the teacher-authors’ conceptions are quite in line with the assessment guidelines. For the content, in 55 papers exact quotations from the guidelines were found. For example, the teacher-author of paper #54 wrote: “[A]ccording to the mathematics curriculum standards, ‘as for assessing students’ mathematics learning, students’ understanding and using mathematics knowledge and skills should be focused on. Students’ development of emotion and attitude also need more attention’.” Although all the competence domains – basic knowledge and skills, mathematical thinking and problem solving, and mathematical and learning attitude – are mentioned in the papers, basic knowledge and skills received most attention from the teacher-authors.

3.3.3 Assessor in assessment

Concerning the assessor, the teachers' conceptions reflected in the papers adhere to the multi-actor assessment system that is promoted in the assessment guidelines. But in the papers, teachers play a dominant role in conducting classroom assessment. Nevertheless, evidence and examples of self- and peer-assessment and parents as assessors can also be found. In paper #64, the teacher-author described: "[T]he [2001] mathematics curriculum standards document points out when assessing students' learning, self- and peer-assessment should be organized."

3.3.4 Method of assessment

Most of the assessment methods recommended in the assessment guidelines in the mathematics curriculum standards were also found in the papers. This particularly applies to the exercises in class and after class ("homework" in the papers) and to the oral tests ("questioning" in the papers). To a lesser degree the papers mentioned doing observations, keeping portfolios, and administering written tests ("quizzes" in the papers). However, while the importance of written tests is emphasized in the mathematics curriculum standards, its use was only found in 25 papers.

3.3.5 Report and use of assessment results

The only aspect of assessment for which we found a real difference between what the mathematics curriculum standards document intends and what is in the papers is the report and use of assessment results. Although in the papers much attention is paid to the report of assessment results, that is, to the provision of feedback, which echoes the guidelines, few discussions focus on the use of classroom assessment results for adapting and improving teachers' further instruction to meet the students' needs.

4. Conclusions and discussion

4.1 Teachers' conceptions of classroom assessment

Through their papers the teacher-authors gave a rich picture about their conceptions of classroom assessment (Research question 1). They described various aspects of their classroom assessment practice, including the purpose, content, assessor, method, and feedback, and illustrated these aspects with examples. Moreover, they gave many reflections on how they assess their students. However, though many teacher-authors explicitly stated that improving their instruction is one of the purposes of conducting classroom assessment, in only a few papers discussions were found about instructional decision-making based on the information gained by their classroom assessment activities. One possible reason for this might be that teachers find it difficult to use assessment-based information for adapting further instruction, as it was found by, for example, Heritage, Kim, Vendlinski, and Herman (2009). A second possible reason could be related to the teacher-authors' narrow scope of classroom assessment. Many of them seem to consider classroom assessment equivalent to providing feedback. This view on assessment is, for example, revealed in some papers in which the teacher-authors either defined classroom assessment as providing students feedback or explained their classroom assessment practice by just reflecting on how they reacted or will react when encountering certain student performances or facing students with certain characteristics. Giving such explanations indicates that the teacher-authors thought their practice of classroom assessment started only after their students' performance. In fact, for some of these teachers, classroom assessment is what a teacher says or writes, in such a way that it helps their students to improve. This is also evidenced by their use of the terms "classroom assessment" and "classroom assessment language" interchangeably, implying that for these teachers the language they use when assessing actually is the assessment.

Undoubtedly, providing feedback to facilitate students to move forward is one of the key strategies for effectively implementing classroom assessment (Black & Wiliam, 2009), and the type of feedback and the way it is given matter its effectiveness of enhancing students' learning (Hattie & Timperley, 2007). From this perspective, it is encouraging to find that how to give feedback was widely discussed by the teacher-authors and that the feedback they gave was mainly task- and process-related, balanced, and timely. Nevertheless, we were surprised by the huge amount of attention that is paid to feedback, especially when comparing this with the attention that is paid to teaching adaptation. Teachers' competence to adapt their instruction based on evidence is critical for effective teaching. If teachers cannot make instructional decisions according to the assessment information, the promise of classroom assessment to improve students' learning will be impaired (Heritage et al., 2009).

In addition, it is interesting that some teacher-authors referred to the beauty aspect of classroom assessment. In fact, in the body of assessment literature known to us and mostly based on studies carried out in the Western world, we never came across references to the aesthetic appeal of assessment. It could be that this approach is typical for how teachers in China conceive education and assessment, but we did not find any substantial evidence for this in other studies. Another possible reason for emphasizing the art and beauty aspect of assessment might be that the teacher-authors would give their paper a good reception by making the topic of assessment more attractive for readers.

4.2 Relation between teachers' conceptions and the assessment guidelines

Regarding the relation between the classroom assessment as reflected in the teacher-written papers and the assessment guidelines in the mathematics curriculum standards (Research question 2), it is evident that the latter has exerted a great influence on inviting primary school mathematics teachers to think about, to discuss, and to share their use of classroom assessment. From 2002 on, one year after the curriculum reform outline (MoE, 2001a) and the mathematics curriculum standards (MoE, 2001b) were published, a steady increase in the number of published papers on classroom assessment was found. In 2011 and 2012, a decade after the advent of the assessment reform, a great number of 266 published teacher-written papers were found to reflect on different aspects of classroom assessment included in the assessment guidelines. By analyzing the papers, we found that the teacher-authors clearly considered these guidelines as a source for conducting their classroom assessment. The teachers often used literal quotations from the documents or paraphrased the guidelines when discussing the purpose of classroom assessment, the content, the assessor, the method and the report of assessment results, that is, giving feedback. The only mismatch with the assessment guidelines lies in the teachers' use of the assessment results. While it is suggested in the mathematics curriculum standards that the assessment results can be used for adapting and improving instruction, this aspect of assessment was hardly addressed in the teacher-written papers. Maybe the reason for this is that the document of the mathematics curriculum standards only briefly describes the use of assessment results.

A further difference between the classroom assessment discussed in the papers and what is suggested in the mathematics curriculum standards relates to the use of written tests or quizzes. Although the guidelines in the latest version of the mathematics curriculum standards (MoE, 2011) state that written tests are important and give detailed suggestions about designing and conducting these tests, only 25 papers mentioned this method. A possible explanation for this discrepancy is that the 2001 version (MoE, 2001b) was rather reluctant in emphasizing the use of written tests in primary education. Taking into account that the reviewed papers were published in 2011 or 2012, it is understandable that the revaluation of written tests in the latest version (MoE, 2011) is not already reflected in the teachers' papers.

When looking back at our analysis and results, indeed we found that the teachers' conceptions of classroom assessment reported in their papers generally were in agreement with the assessment guidelines. However, together with this positive finding, in retrospect we have noticed that the assessment guidelines also may have affected the teachers' conception of classroom assessment in another way. Although the structure of the assessment guidelines in the Chinese mathematics curriculum standards is very clear by addressing all the key aspects of performing classroom assessment specifically and providing suggestions or examples for each aspect, this structured presentation also can have a disadvantage. By delivering the message about the new approach to assessment aspect-by-aspect, it becomes maybe not so clear for teachers how these aspects of classroom assessment function as a whole and how to embed them in their daily teaching practice. Therefore, this structure may hinder teachers to get a holistic picture of how classroom assessment works – which is reflected in the teachers' papers – and this might jeopardize the implementation of classroom assessment.

Possible improvements regarding to the presentation of the assessment guidelines can be put on the agenda of the policy makers of the Chinese mathematics curriculum standards. For example, detailed suggestions about how teachers can make use of the assessment results to adapt and improve their teaching are needed. Moreover, after addressing all the key aspects of assessment separately, explanations about how these key aspects work as a whole and how to integrate assessment into teachers' practice should be provided. Finally, it may be helpful to give teachers concrete examples of conducting assessment, which illustrate, in the context of assessing specific mathematics content, for what purposes certain assessment methods are chosen to use, possible assessment results teachers and students may receive, diverse feedback that teachers can provide to students, and various instructional adaptations teachers can do for their further teaching.

4.3 Limitations and recommendations for future research

As our search in the CNKI database has shown, assessment at the classroom level in primary mathematics education is a rather recent educational phenomenon in China and is rapidly gaining ground. How this reform in assessment took place and is still going on, and what were or are the important change agents in getting this reform implemented in educational practice would be very relevant information for further improving education. However, due to the limited time for carrying out this review study we could only include the papers published in 2011 and 2012. To gain a complete picture of the change in classroom assessment, in future research more years could be included. Another extension that is recommended is to have a closer look at the papers written by researchers and consultants, whose views were left out in our review due to the very small number of such papers that were found in the database.

Finally, although teacher-written papers are a valuable source for getting more knowledge about teachers' conceptions of classroom assessment, the findings from the 266 teacher-written papers need to be interpreted with prudence, since only teachers who had their papers on classroom assessment published were included in this study and the teachers' reports for publication may have been affected by social desirability and publication bias. Another concern is that "there are profound gaps between what people know, what they think they know, what they say, and what they do."[1] Therefore, it is unsure whether what is written in the teacher-written papers can be considered fully equivalent to what the teachers really think of assessment. Further sources could include directly asking teachers about their classroom assessment or carrying out classroom observations to describe what appears to be happening in classrooms. Conducting a survey to further investigate Chinese primary mathematics teachers' use of assessment will be the next step in our research.

[1] We thank an anonymous reviewer for suggesting this phrasing.

References

Andrade, H. L. (2010). Students as the definitive source of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). Abingdon, UK: Routledge.

Assessment Reform Group (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge School of Education.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Berry, R. (2011). Educational assessment in mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 49–61). Netherlands: Springer.

Black, P., & Wiliam, D. (1998a). Inside the black box. Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

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

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

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.

Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429–458.

Cao, Y., Li, J., & Qing, H. (2011). Woguo shuxue ketang jiaoxue pingjia yanjiu zongshu [A review of classroom instruction and learning assessment in mathematics education in China]. *Mathematics Bulletin, 8*, 1–5.

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.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

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.

Herse, R. (1979). Some proposals for reviving the philosophy of mathematics. *Advances in Mathematics, 31*(1), 31–50.

Jianli. (1949). Jianli yishi baogao yu pingding wujiao zhidu de tongzhi [Announcement about reporting and evaluating the price of materials instantly]. *Governmental Journal of Jiangxi, 2*, 55–60.

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.

Leung, F. K. (2008). In the books there are golden houses: Mathematics assessment in East Asia. *The International Journal on Mathematics Education, 40*(6), 983–992.

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

Ministry of Education of the People's Republic of China (MoE). (2001a). *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). (2001b). *Quanrizhi yiwu jiaoyu shuxue kecheng biaozhun (shiyangao)* [Mathematics curriculum standards of nine-year compulsory education (trial version)]. Beijing, China: Beijing Normal University Press.

Ministry of Education of the People's Republic of China (MoE). (2007). *Quanrizhi yiwu jiaoyu shuxue kecheng biaozhun (xiugaigao)* [Mathematics curriculum standards of nine-year compulsory education (revised version)]. Retrieved from http://wenku.baidu.com/view/3b7a694233687e21af45a93e.html

Ministry of Education of the People's Republic of China (MoE). (2011). *Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (NCTM). (2013). *Formative Assessment: A position of the National Council of Teachers of Mathematics.* National Council of Teachers of Mathematics. Retrieved from http://www.nctm.org/about/content.aspx?id=37990

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: Emerald.

Romberg, T. A., Zarinnia, E. A., & Collis, K. F. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 21–38). Washington, DC: AAAS.

Rong, Z. (1985) Xuexi zhiliang fenxi kapian de sheji yu shiyong [The design and use "the worksheet for analyzing the quality of learng": Management of teaching and learning in primary mathematics education]. *Theory and Practice of Education, 5*, 23–26.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Shi, N., Ma, Y., & Liu, X. (2012) Yiwu jiaoyu shuxue kecheng biaozhun xiuding guocheng yu zhuyao neirong [The revision process and the content revised in the mathematics curriculum standards of nine-year compulsory education]. *Curriculum, Teaching Material and Method, 32*(3), 50–56.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11–18.

Stiggins, R. (2002) Assessment crisis: The absence of Assessment For Learning. *Phi Delta Kappan, 83*(10), 758–765.

Suurtamm, C., Koch, M., & Arden, A. (2010) Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399–417.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15*(2), 105–127.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York, US: Macmillan.

Topping, K. J. (2010). Peers as a source of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 61–74). Abingdon, UK: Routledge.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education*. Utrecht, the Netherlands: CD Beta Press.

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.

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 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.

Wang, S., Yuan, L., Tian, Z., & Zhang, X. (2013). Woguo ge shengfen jiaoyu fazhan shuiping bijiao fenxi [The Comparative Research on the level of regional development of education]. *Research in Education, 6*, 29–41.

Wang, X. (2011). Ketang pingjia ying guanzhu xuesheng de zhuangtai [Classroom assessment should focus on students' situation]. *Education of Liaoning, 11*, 45–46.

Wiliam, 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.

Wiliam, D. (2011a). *Embedded Formative Assessment*. Bloomington, IN: Solution Tree.

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

Zhong, Q. (2012). Ketang pingjia de tiaozhan [The challenges of classroom evaluation]. *Global Education, 1*(10), 10–16.

Zhu, L. (2012). Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban) xiuding le shenme [What has the mathematics curriculum standards of nine-year compulsory education (2011 version) revised]. *Journal of Mathematics Education, 21*(3), 7–10.

Authors' contributions

This paper was a collaborative work of the three authors. XZ, MHP, and MV all participated in designing the study. XZ carried out the data collection in the CNKI database, and was responsible for interpreting the Chinese resources. XZ, MHP, and MV analyzed the data, and drafted and revised the manuscript. All authors read and approved the final manuscript.

Chapter 3

Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey

Xiaoyan Zhao[a], Marja van den Heuvel-Panhuizen[a,b], & Michiel Veldhuis[b]
[a]*Freudenthal Institute, Faculty of Science, Utrecht University, the Netherlands*
[b]*Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2017). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education*, Advance online publication. doi:10.1007/s10763-017-9841-3

Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey

Abstract

This study investigated Chinese primary school mathematics teachers' views on assessment in an effort to determine their assessment profiles. A large-scale questionnaire survey with 1101 teachers from 12 Chinese provinces and regions was carried out. The teachers reported to use assessment on a daily or weekly basis for different purposes. They recognized the importance of assessing different types of skills and knowledge and considered assessment useful for improving teaching and learning. To determine teachers' assessment profiles, we used several latent variable modeling techniques. With exploratory factor analyses, we identified eight factors in the teachers' responses: general instructional decision-making assessment purposes [1], specific instructional decision-making assessment purposes [2], assessment methods [3], diversity of assessment problem format [4], importance of assessing skills and knowledge [5], importance of assessing extra-curricular skills [6], perceived usefulness [7], and acceptance of assessment [8]. When these factors were used to interpret the results of a latent class analysis, three distinct assessment profiles could be distinguished. One fifth of the teachers were in the *Enthusiastic assessors* profile. These teachers not only reported to use assessment frequently [3, 4] and purposefully [1, 2], but also highly endorsed its importance [5, 6] and usefulness [7, 8]. Around half of the teachers were in the *Mainstream assessors* profile; these teachers scored close to the mean on all factors. The remaining teachers held the relatively negative views on assessment and were therefore in the *Unenthusiastic assessors* profile. This profile characterization sheds light on Chinese primary school mathematics teachers' assessment culture.

Keywords: Assessment; mainland China; mathematics education; primary school; survey

1. Introduction

Assessment is crucial for teaching and learning at all educational levels and across all school subjects. Without assessment, it is hard to determine whether students have achieved the intended goals and to make instructional decisions about how students can best be helped to reach these goals. This latter purpose of assessment, which focuses on supporting the students' further learning, has gotten more attention over the last 20 years. Awareness has arisen that assessment should not only serve summative purposes, for example, using it for grading students, but should place more emphasis on formative purposes, such as informing teachers' instruction and improving students' learning (e.g., Assessment Reform Group, 1999; Black & Wiliam, 1998; Stiggins, 2002). Another change in the assessment policy and practice is that assessment is increasingly put in the hands of the teachers, because they are considered to be in a good position for collecting information about their students' learning (Harlen, 2007). This means that assessment is interwoven with instruction as an on-going process, which offers teachers direct information to make adequate instructional decisions, in order to cater their students' needs and, in this way, can raise the achievements of their students. Several studies have evidenced this power of teachers' assessment activities to improve students' mathematics learning (e.g., Cauley & McMillan, 2010; Phelan, Choi, Vendlinski, Baker, & Herman, 2011; Veldhuis & Van den Heuvel-Panhuizen, 2014a; Wiliam, Lee, Harrison, & Black, 2004). As a result of these promising findings, the teachers' assessment practice has become a key factor in improving mathematics education and has been put on the policy agendas in many countries (Berry, 2011).

In line with this, investigations have been carried out all over the world to find out mathematics teachers' current assessment practice and beliefs on assessment. The present study is meant to do such an investigation in China to gain knowledge about how primary school mathematics teachers in China consider and perform assessment in their teaching.

2. Literature review

2.1 Teachers' assessment practice and beliefs

Through surveys based on interviews (e.g., in Finland: Krzywacki, Koistinen, & Lavonen, 2011; in the USA: Riggan & Oláh, 2011; in Canada: Suurtamm, Koch, & Arden, 2010) and questionnaires (e.g., in Canada: Suurtamm et al., 2010; in China: Ni, Li, Li, & Zou, 2011; in the Netherlands: Veldhuis et al., 2013), it was found that teachers reported to use various assessment methods. Particularly, it seems that teachers tend to use observation-based assessment methods, like questioning, observing, and correcting written work, for formative purposes. At the same time, it was also found that teachers rely on instrument-based methods, like paper-and-pencil tests, for summative purposes (Riggan & Oláh, 2011; Suurtamm et al., 2010; Veldhuis et al., 2013). These findings were not only extracted from teachers' self-reported data, but also confirmed by classroom observations (Riggan & Oláh, 2011; Suurtamm et al., 2010). However, not in all countries the assessment practice of teachers is well established. For example, when reviewing policy documents and research reports from Norway and Portugal, Nortvedt, Santos, and Pinto (2016) found that in these countries the intended assessment is only scarcely implemented in primary mathematics education. Furthermore, in an online questionnaire study conducted in the USA to measure teachers' assessment proficiency (Heritage, Kim, Vendlinski, & Herman, 2009), it was uncovered that teachers have difficulties in using assessment information to decide their next teaching steps came to the fore.

Regarding the beliefs of teachers on assessment, several large-scale questionnaire survey studies done in several countries by Brown and his colleagues (e.g. in Australia: Brown, Lake, & Matters, 2011; in New Zealand: Brown, 2004; in India: Brown, Chaudhry, & Dhamija, 2015; in China: Brown, Kennedy, Fok, Chan, & Yu, 2009; Chen & Brown, 2016) revealed that teachers in general, mathematics teachers included, tend to embrace the idea of using assessment to improve teachers' instruction and students' learning by the provision of quality information for making instructional decisions. Furthermore, a later study carried out in New Zealand by Brown (2009) showed that having improvement-orientated assessment beliefs can predict teachers' increased assessment practice. Yet, holding particular beliefs on assessment is no guarantee for a corresponding assessment practice. As was shown in a large-scale questionnaire survey in the UK by Sach (2012), despite that the teachers clearly acknowledge the value of formative assessment in promoting learning, their responses suggested that they are less confident than they claim to be in implementing the assessment strategies in their classroom practice.

A further step in researching how assessment is conceptualized and operationalized by teachers is to identify particular characterizations of the teachers' views on assessment. This is an approach that can lead to different groups of teachers whose perceived assessment practice and assessment beliefs are each based on particular combinations of their responses regarding various aspects of assessment. An example of this approach is worked out by Veldhuis and Van den Heuvel-Panhuizen (2014b) who identified four different assessment profiles of teachers –consisting of *Enthusiastic*, *Mainstream*, *Non-enthusiastic*, and *Alternative assessors*– based on data collected by an online questionnaire in a sample of teachers in the Netherlands. Another recent example is the study carried out by Barnes, Fives, and Dacey (2017) in the USA. They identified three distinct profiles in terms of teachers' conceptions of assessment purposes, based on teachers' perception of the relevance of assessment, its validity for accountability, and its use to improve teaching and learning. Also for the general teaching skills of teachers, different types of teacher behavior have been identified. For example, Kyriakides, Creemers, and Antoniou (2009) found five types ranging from *Basic elements of direct teaching* to *Achieving quality and differentiation in teaching using different approaches*. According to the authors, these types of teacher behavior could be interpreted as stage models of professional development and were considered as relevant for supporting professional development of teachers.

2.2 Assessment reform in mathematics education in China

In 2001, in the People's Republic of China, a new curriculum for teaching mathematics was launched by the Ministry of Education (MoE, 2001). Compared to the previous curriculum, more attention was paid to students' mathematical thinking and problemsolving ability, while the traditional merits of emphasizing basic knowledge and skills in mathematics education were still maintained. Students' ownership of their learning was highlighted, and they were encouraged to learn through active participation, cooperation, and communication. At the same time, teachers' roles as organizers, facilitators, and cooperators were also made clear.

Together with this curriculum reform an assessment reform was initiated, which called for reducing the overemphasis on using assessment for selection purposes, and establishing an improvement-oriented assessment system that supports teaching and learning. To better help mathematics teachers in compulsory education, which is from Grade 1 to 9, to put this new idea of assessment into action, guidelines for assessment were published in the mathematics curriculum standards (MoE, 2001, 2011). Particularly, the main purpose of assessment, the content of assessment, the person who is the assessor, the assessment methods, and suitable ways of reporting and using assessment results are discussed. According to the latest version of the mathematics curriculum standards (MoE, 2011),

> the main purpose of assessment is getting the whole picture of process and outcomes of student's mathematics learning, stimulating students to learn, and improving teachers' instruction. (MoE, 2011, p. 52)

For the content of assessment, it is stipulated that assessment should address what mathematics students have to learn and what mathematical competences they have to develop. Advices are provided about how to assess students' basic knowledge and skills, their mathematical thinking and problem solving, and their learning attitude. Regarding the person who is conducting assessment, the assessment guidelines suggest establishing a multi-actor system of assessment, in which not only the mathematics teacher, but also students, their peers, and parents can be involved in the assessment. Moreover, various assessment methods are recommended to be used for getting information about student learning, like written tests, oral tests, open questions, activity reports, observations, interviews, exercises in and after class, and portfolios. Teachers are required to understand the characteristics of different assessment methods, and to be able to choose appropriate methods that fit both the content to be assessed and their students' learning situation. Also, the assessment guidelines refer to reporting and using assessment results. The assessment results should be reported in a way that can enhance students' confidence and learning interests, can help them to develop good learning habits, and can facilitate their learning. Moreover, it is described how teachers can benefit from assessment results by adapting and improving instruction based on information about their students' learning. Although the assessment guidelines in the Chinese mathematics curriculum standards cover all the key aspects of using assessment for the purpose of supporting teaching and learning, the practical suggestions given for each aspect are quite brief.

2.3 Implementation of the assessment reform in primary school mathematics in China

Since 2001, a number of studies have been carried out on the implementation of the assessment reform in practice. These studies focused on all kinds of subjects and mainly in secondary education (e.g., Brown & Gao, 2015; Chen & Brown, 2016). To our knowledge, only scarce attention has been paid in research to whether, and to what extent, the new approach to assessment has been implemented in primary mathematics classroom. One of the studies we found is a case study carried out by Zhao, Mulligan, and Mitchelmore (2006) in which six primary mathematics teachers were observed and interviewed shortly after the start of the assessment reform. This study revealed that, for these teachers, external and formal examinations still play a dominant role in their assessment activities and that the students are not actively involved. These findings suggest that, at that moment, there was still a considerable mismatch between the intended assessment advocated by the assessment guidelines and the investigated teachers' assessment practice. Obviously, and this is also what one might have expected, it takes some time before teachers become familiar with a new approach to assessment. This was shown by a large-scale questionnaire survey that was conducted in 2005 by Ni et al. (2011) in which 390 primary mathematics teachers from Henan province were involved. Based on this survey that was focused on the implementation of the curriculum reform in mathematics education in general, it was found for assessment that 4 years after the launch of the assessment reform, the teachers were able to employ assessment methods as recommended in the assessment guidelines.

Regarding the beliefs on assessment, we found two studies in China. In the case study of Zhao et al. (2006), it was revealed that most of the participating primary mathematics teachers recognize the importance of assessment for improving their teaching. At the same time, however, they believed that the major purpose of assessment is to inspect students' mathematics learning in order to stimulate students' motivation to improve their achievement level. Further information about teachers' beliefs on assessment in mainland China comes from a large-scale survey starting in 2008 that was carried out by Brown, Hui, Yu, and Kennedy (2011). In this study, 898 teachers from Southern China filled in a questionnaire with 30 questions about what they think about the nature and purposes of assessment. The teachers' responses revealed that they highly endorse assessment leading to the improvement of the teaching quality, students' learning and personal development, and that the teachers also value the accountability purpose of assessment. Yet, in this research only 3% of the respondents were mathematics teachers, including both primary and secondary school teachers.

In addition to the studies done by researchers, also papers published by teachers themselves can give evidence of the implementation of the assessment guidelines in classroom practice. Based on a review of 266 teacher-written papers included in the China National Knowledge Infrastructure (CNKI) database and published in the years 2011 and 2012, it was found that primary school mathematics teachers' conception of classroom assessment and their reported assessment practice echo well with the assessment guidelines (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2017, *Chapter 2 of this thesis*). The only point that was just scarcely discussed by the teacher-authors is using assessment information to adapt and improve further instruction. In many of these teacher-written papers, assessment conducted by teachers at classroom level is considered to be equivalent to the provision of feedback.

3. Research question

The aforementioned studies have shed some light on how primary mathematics teachers use and perceive the assessment as advocated in the assessment reform launched in mainland China in 2001. Apart from teachers' assessment practice and beliefs as reflected in teacher-written papers, there are, as far as we know, only three research papers (i.e., Brown, Hui et al., 2011; Ni et al., 2011; Zhao et al., 2006) which provide some information about the implementation of the assessment reform in primary mathematics education in mainland China. The most recent data collected by these three studies date from 2008, which means that little is known about how the implementation of the assessment reform has further evolved. So, one may conclude that knowledge about primary mathematics teachers' current assessment practice and beliefs is in need of an update. Also research is necessary which has a broader scope than the previous studies, both in the number of teachers involved and the regions of mainland China covered. Therefore, we set up the current study. In order to gather information from primary school mathematics teachers from all over mainland China, we chose for a large-scale survey based on a written questionnaire. Our main research question was *What assessment profiles can be identified in Chinese primary school mathematics teachers?*

4. Method

To answer our research question we first looked into how Chinese mathematics teachers in primary education view their assessment practice. This means that we questioned the teachers about all aspects related to how they assess their students and how they think about assessment. In addition to this specific information about what teachers do in their classrooms in the name of assessment and what their beliefs are on assessment, we aimed to obtain a more general picture about the presence of particular assessment cultures. Specifically, we investigated whether it is possible to distinguish groups of teachers, for which the views differ between groups, but are similar within each group.

4.1 Instrument

For developing the questionnaire for this survey we made use of a questionnaire used in the Netherlands for investigating the teachers' assessment practice and beliefs (Veldhuis & Van den Heuvel-Panhuizen, 2014b; Veldhuis et al., 2013). The original Dutch questionnaire contained 40 questions by which data could be collected about primary school teachers' mathematics teaching practice, their assessment practice, and their beliefs on assessment, and some personal and professional background information. The questions were generally based on literature about assessment. The possible assessment methods and purposes were deduced from Black and Wiliam (1998), Stiggins and Bridgeford (1985), Mavrommatis (1997), and Suurtamm et al. (2010). The questions aimed at investigating teachers' beliefs on assessment were adapted from Brown's (2004) Teachers' Conception of Assessment (COA-III) questionnaire.

When adjusting the Dutch questionnaire for using it with Chinese teachers, some questions or items in the Dutch questionnaire were deleted or adapted, because they did not fit to the Chinese situation. For example, in the Chinese version, no questions were asked about standardized tests at district or city level, because such tests are not generally used in all Chinese regions (Pan, 2015). Another adaptation was that we extended the six-point scale to a seven-point scale by including "daily" when teachers have to indicate the frequency of their assessment practice. The reason for this was that Chinese primary mathematics teachers normally plan and give mathematics lessons on a daily basis according to a fixed school timetable. Before the questionnaire was used in our study, it was piloted. A first version of the adapted questionnaire was filled in by 18 primary mathematics teachers from four schools in different provinces in China; their comments were used to improve the questionnaire by adding further clarifications and changing the wording of the questions.

The final version of the questionnaire consisted of 30 questions. The first 10 questions were aimed at collecting teachers' background information, such as their age, gender, educational background, and teaching experience. The next 12 questions were used to characterize mathematics teachers' general teaching practice. Among other things, information was gathered about whether teachers divide their students into different level groups, whether they discuss students' learning with other colleagues, and whether also students, parents, and other staff in school are involved in assessment.

The remaining eight questions were focused on how teachers view their assessment practice. Specifically, to investigate for what purposes and by which methods teachers assess their students, two series of questions were provided and teachers needed to rate on a seven-point-scale how often they carry out possible assessment purposes and methods (1 = Rarely to never, 2 = Yearly, 3 = A few times a year, 4 = Monthly, 5 = Weekly, 6 = A few times a week, 7 = Daily). For example, teachers were asked to tick how often they use assessment with the aim to determine students' mastery of certain mathematics topics, to provide feedback to students, or to formulate learning goals; and how often they assess students by means of asking questions, keeping portfolios, or using textbook tests. Furthermore, we asked the teachers to indicate the types of exercises they used for assessing their students, for example, bare number problems, problems in context, and problems having multiple solutions. In all these questions about purposes, methods, and types of problems, teachers were given the opportunity to extend the possibilities listed in the questionnaire. The next series of questions addressed the perceived importance of the assessment content. Teachers were required to rate the importance of assessing particular knowledge and skills on a four-point-scale (1 = Very unimportant, 2 = Unimportant, 3 = Important, 4 = Very important). Finally, teachers were invited to indicate their agreement with a series of statements about assessment on a four-point-scale (1 = Completely disagree, 2 = Disagree, 3 = Agree, 4 = Completely agree). Two examples of these statements are "assessment is not influencing my teaching" and "assessment is useful for helping students to learn."

4.2 Data collection

Data collection was carried out from the end of February to the end of April, 2013. As the educational situation varies largely between provinces and regions in China, we decided to collect data in as many different places as possible. In practice, we contacted volunteers from the first authors' circle of acquaintances from different places in China to assist us in our study. The volunteers were former classmates who are now teachers in primary school or educational consultants in a district. These volunteers were responsible for printing the questionnaires, handing them out to primary mathematics teachers, and explaining the purpose of the survey.

4.3 Sample

In total, the questionnaire was returned by 1172 primary mathematics teachers. However, some questionnaires could not be used because no question about assessment was answered by the teachers. Also, some questionnaires were lost in the process. This resulted in a final sample of 1101 primary mathematics teachers whose questionnaires we could use in the analysis. The teachers involved were from 12 out of the 31 provinces, municipalities, and autonomous regions in mainland China. Half of the teachers were from Hebei province, where the overall level of educational development is above the average; the educational development of Hebei is ranked 13th out of 31 provinces, municipalities, and autonomous regions (Wang, Yuan, Tian, & Zhang, 2013). One fifth of the teachers involved were from Jiangsu province, which is ranked in the 5th place of educational development (Wang et al., 2013). The remaining teachers (29% of the total sample) were from 10 other provinces or municipalities.

4.4 Data analysis

Before we started the analyses, we checked the inputted data and cleaned them where necessary. Some teachers appeared to have given illogical answers, for example one teacher said her age was four. Such answers were recoded as missing. Also, we detected some clear coding mistakes, where answers were put into incorrect columns for example. In these cases, we corrected the coding. We started with analyzing the factorial structure of the questionnaire and report descriptive statistics on the teachers' reported general teaching and assessment practice. Then, latent class analysis was used to determine these Chinese teachers' assessment profiles (Veldhuis & Van den Heuvel-Panhuizen, 2014b).

The factor analysis was based on the answers to the eight questions that focused on how teachers view their assessment practice. To identify the underlying latent structure of the items in the questionnaire we employed several latent variable modeling techniques. To decide about the most appropriate model we used substantive as well as statistical model fit checking (Muthén, 2003). For our substantive model checking, we checked whether the model's predictions and constituents were in line with theoretical and practical expectations. To evaluate the statistical model-data fit we checked, for the factor analyses, the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and a chi-square statistic (Barrett, 2007). We used the conventions for acceptable model fit of RMSEA below 0.06 and the CFI over 0.96 (Hu & Bentler, 1999). In these factor analyses, we first envisioned a confirmatory approach, as our questionnaire was based on an existing instrument, however, the confirmatory model replicating the Dutch latent structure did not reach convergence. Therefore, we proceeded with performing a number of exploratory factor analyses with weighted least squares method (WLSM) estimation and geomin oblique rotation to determine the structure of variation on the measured variables. When models reached convergence and had satisfactory fit indices, we checked whether the factors made substantive sense and looked if the items making up the factors had sufficiently in common and allowed us to name them accordingly. To decide upon the best fitting model, we combined the results of the substantive and the statistical arguments.

In parallel, we performed latent class analyses to identify underlying classes of teachers based on differences in the patterns of their responses on items in the questionnaire. To decide upon the number of classes, we looked at the Bayesian Information Criterion (BIC), the relatively lowest value indicates the best fit, and entropy (Dias & Vermunt, 2006). The teachers were assigned to a latent class –that we will call assessment profiles– through modal assignment, i.e., they were assigned to the latent class to which they had the highest probability of belonging.

Finally, differences between teachers with the different assessment profiles on a number of background variables were investigated with analyses of variance (ANOVA), Kruskal-Wallis, and χ^2-differences tests. With these analyses, the defining elements for each profile could be determined. The inferential analyses were performed in SPSS 23 (IBM Corp, 2014) and all latent variable modeling in MPlus 6 (Muthén & Muthén, 1998–2010).

5. Results

5.1 Teachers' characteristics and their general teaching practice

The teachers in the final sample were mostly female (85%). Their mean age was 36.0 years (*SD* = 7.2), with average teaching experience of 13.3 years (*SD* = 8.3). Around a quarter (26%) of the teachers had worked for 1 to 6 years, another quarter (26%) for 7 to 13 years, the next quarter (25%) for 13 to 19 years, and the last quarter (25%) for 20 years or more. Most of the teachers (93%) were educated to become a teacher. A few teachers (3%) only graduated from secondary school; some (37%) graduated from technical secondary school; some (30%) had an associate bachelor's degree; and some others (28%) had a bachelor's degree. Only 23 teachers (2%) had a master's degree. The sample in our study covered only a small proportion of the large population of the about 1.7 million primary school mathematics teachers that China had in 2013 (MoE, 2014). Compared to the whole population (57% female teachers), we had proportionally more female teachers in our sample. With respect to the teachers' educational background our sample has about the same proportion of primary school mathematics teachers with a Master's degree and a Bachelor's degree as were in the whole population.

The participating teachers taught students in different grades. Except for ten teachers who reported to teach kindergarten children, most of the teachers taught Grade 5 (20%) and Grade 6 (20%), the least teachers (15%) taught Grade 3. More than half of the teachers (63%) taught only one class. If the teachers had more classes, nearly all ofthem (94%) taught students in one grade. The average class size was 54 (*SD* = 16), which differs from the national average of about 37 students (MoE, 2014; OECD, 2012, p. 450).

Of the 1018 teachers who responded to the question whether they received professional development in 2012 –which is the year before the study was carried out– a few (13%) reported that they did not attend any professional development meeting. More than half of the teachers (56%) wrote that they participated in up to three meetings; the remaining teachers (31%) mentioned that they had trainings for more than three times. The themes of the professional development meetings were also provided by 750 teachers: the comprehension of the new mathematics curriculum standards was mentioned most (27%), followed by the use of textbooks (12%). Only five teachers explicitly referred to "assessment", in Chinese *PingJia* (评价); 29 teachers provided topics related to assessment, like how to pose questions or how to deal with students' mistakes.

Most teachers (90%) reported to give mathematics lessons every day. According to the teachers' report, the main focus in these lessons was on giving instruction ($M = 42\%$, $SD = 0.15$) or on asking students to finish exercises ($M = 41\%$, $SD = 0.16$), whereas lesser time was reserved for assessing students ($M = 15\%$, $SD = 0.08$). The vast majority of the teachers (92%) answered that they have clear and specific goals for their students' mathematics learning. A few teachers (11%) stated that they almost never share the learning goals with their students; more teachers reported to share the goals monthly (27%) or weekly (24%); a small number of teachers (6%) responded to share the goals daily. Regarding having level groups in class, a few teachers (7%) answered that they do not distinguish different level groups; the majority (82%) wrote that they make a distinction between students with different capabilities, but only in their mind; the remaining teachers (27%) mentioned to organize their classroom in a way that students of the same level sit together. In addition, most teachers (98%) reported that they discuss their students' learning with either the teacher who is responsible for the general management of the class or the teachers who teach other subjects in the class. More than half of the teachers mentioned that students themselves (65%) and their peers (50%) are involved as assessors. A few teachers also referred to someone from the school management department (11%) or students' parents (10%) as assessors.

5.2 Teachers' assessment views

After comparing one- to eight-factor solutions, our exploratory factor analyses delivered an eight-factor solution that had a good enough fit ($\chi^2(938, N = 1076) = 3030.5$, $p < .0001$, RMSEA = .045, CFI = .97). Also, these eight factors all had eigenvalues over 1.5. The χ^2 statistic of the overall model fit was significant, which indicates a model with a less than optimal fit. Nevertheless, this nested eight-factor solution fitted significantly better than the seven-factor solution, as illustrated by the Satorra-Bentler scaled χ^2 test, which is unaffected by non-normality (TRd (df = 45) = 416.5, $p < .0001$). Most of the subscales in the questionnaire loaded coherently on different latent factors providing substantive evidence for this eight-factor solution (see Tables 1–8 for the items constituting the latent factors and the corresponding scale's Cronbach's alpha).

Taking into account the content of the items making up the eight factors, we decided on the following names: (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 of assessment*, and (8) *Acceptance of assessment*.

In the factor *General instructional decision-making assessment purposes* (Table 1) were those items of the subscale on the purposes of assessment being related to more general instructional decision-making by the teacher, such as, determining students' mastery or the formulation of learning goals. In the factor *Specific instructional decision-making assessment purposes* (Table 2) were items that were more related to specific instructional decision-making, such as investigating reasons for student errors or stimulating students to think about their solutions. Most participating teachers reported that they use assessment for the different purposes on a daily or weekly basis (> 63%). On a daily basis, stimulating students' use of scrap paper (74%) was mentioned most, followed by stimulating students to think about their solutions (62%). Concerning these two factors on the purposes of assessment, the teachers generally reported to use assessment more frequently for the purpose of making specific instructional decisions (> 90%) than general instructional decisions (> 63%).

Table 1
Factor loadings of the items on General instructional decision-making assessment purposes (α = 0.812)

General instructional decision-making assessment purpose	Factor loading
Determine mastery	0.861
Formulate learning goals	0.713
Determine progress	0.704
Adapt instruction	0.525
Determine speed of instruction	0.406

Table 2
Factor loadings of the items on Specific instructional decision-making assessment purposes (α = 0.851)

Specific instructional decision-making assessment purpose	Factor loading
Select mathematics topics	0.722
Investigate reasons for students' errors	0.714
Stimulate students to think about solutions	0.640
Refer students to further care	0.603
Stimulate students' use of scrap paper	0.554
Provide students with feedback	0.493

The factor of *Assessment methods* (Table 3) was completely made up of the items in the subscale about teachers' assessment methods. Most of the teachers reported that, every day, they assess their students by asking questions (91%), correcting written work (90%), using textbook test problems (78%), and observing (73%). In addition, the majority of the teachers replied that they assessed their students on a weekly basis by using student-development test problems (59%), assigning practical work (47%), asking students to give presentation (46%), and collecting students' scrap paper (43%).

Table 3
Factor loadings of the items on Assessment methods (α = 0.677)

Assessment method	Factor loading
Questioning	0.823
Correcting written work	0.695
Textbook test problems	0.567
Observation	0.506
Student-developed test problems	0.497
Presentation	0.468
Practical assignments	0.452
Collecting scrap paper	0.375

The *Diversity of assessment problem format* factor (Table 4) consisted of the items on the type of mathematics exercises teachers included in mathematics tests. Mathematical problems in context (77%) were used by most of the teachers, followed by variation problems (67%) and mathematical problems with more than one correct answer (65%). Bare mathematical problems (45%) were used the least often by the teachers.

Table 4
Factor loadings of the items on Diversity of assessment problem format (α = 0.699)

Assessment problem format	Factor loading
Mathematical problems in context	0.857
Mathematical problems with more than one correct answer	0.725
Variation problems	0.664
Mathematical problems having multiple solutions	0.647
Practical mathematical problems	0.578
Mathematical problems asking for students' explanation	0.539
Bare mathematical problems	0.513

The items on the importance of assessing different types of skills and knowledge were made up the factor of *Importance of assessing skills and knowledge* (Table 5). For all kinds of knowledge or skills, more than 90% of the teachers reported that they are important or very important to be assessed. A subset of these items, namely, assessing students' evaluation and design skills, made up the factor *Importance of assessing extra-curricular skills* (Table 6). These skills were named as extra-curricular skills because they are barely included in the mathematics curriculum standards (MoE, 2011).

The factor of *Perceived usefulness of assessment* (Table 7) comprised the items with statements about assessment such as assessment helps students to learn. The majority of the teachers indicated that they agreed with the statements. Particularly, 99% of the teachers confirmed that assessment is useful to help students' learning, and 97% of the teachers thought of assessment as useful to improve their instruction. Yet also 40% of the teachers indicated their disagreement with assessment to predict students' performances.

Table 5
Factor loadings of the items on Importance of assessing skills and knowledge (α = 0.823)

Skill or knowledge important to be assessed	Factor loading
Skills for doing analyses	0.799
Skills related to understanding	0.724
Evaluation skills	0.722
Application skills	0.707
Design skills	0.657
Memory skills	0.655
Conceptual knowledge	0.642
Factual knowledge	0.547
Procedural knowledge	0.539
Self-knowledge	0.445

Table 6
Factor loadings of the items on Importance of assessing extra-curricular skills (α = 0.691)

Extra-curricular skill important to be assessed	Factor loading
Design skills	0.527
Evaluation skills	0.363

Table 7
Factor loadings of the items on Perceived usefulness of assessment (α = 0.794)

Usefulness of assessment	Factor loading
Helps students to learn	0.783
Helps to improve teaching	0.740
Provides information about learning needs	0.712
Predicts students' performances	0.676
Creates a better learning climate	0.674
Discloses what students have learned	0.650
Reveals students' strong/weak sides	0.602

Finally, the factor *Acceptance of assessment* (Table 8) consisted of items through which agreement is expressed with the statements that the assessment does not interrupt the teacher's teaching and has much influence on this teaching, together with items that refer to the usual assessment methods of questioning and correcting written work. Most teachers agreed with these statements, however, about one third of the teachers (35%) stated that assessment actually has no influence on their teaching; some others (15%) even mentioned that assessment interrupts their teaching. A small part of the teachers (21%) considered that assessment does not tell them what their students can do.

Table 8
Factor loadings of the items on Acceptance of assessment (α = 0.701)

Acceptance of assessment	Factor loading
Assessment tells me what students can do[a]	0.851
Assessment does not interrupt my teaching[a]	0.836
Assessment has much influence on my teaching[a]	0.575
Assessment method: Questioning	0.418
Assessment method: Correcting written work	0.380

[a] These statements were originally phrased negatively in the questionnaire, e.g., "Assessment does *not* tell me what students can do", and have been recoded

Correlations between these eight factors are displayed in Table 9. The factors related to teachers' assessment practice, namely, the two types of assessment purposes and the assessment methods correlate relatively highly ($.45 < r < .55$). Also the two factors on the importance of assessing skills and knowledge, and extra-curricular skills correlate highly with each other ($r = .70$). The factor of Diversity of assessment problem format stands out in the sense that it only has low or non-significant correlations with the other factors. Looking more closely at the other correlations reveals that the remaining factors correlate weakly to moderately positively with each other ($.092 < r < .400$).

Table 9
Correlations among the eight factors from the exploratory factor analysis (Ns > 1060)

Factors	GAP	SAP	AM	DAF	IASK	IAECS	PUA	AA
GAP (General assessment purposes)	-							
SAP (Specific assessment purposes)	.549	-						
AM (Assessment methods)	.453	.474	-					
DAF (Diversity of problem format)	.050 (n.s.)	.120	.092	-				
IASK (Importance skills and knowledge)	.283	.297	.357	.122	-			
IAECS (Importance extra-curricular skills)	.235	.233	.320	.098	.700	-		
PUA (Perceived usefulness of assessment)	.294	.220	.248	-.010 (n.s.)	.400	.299	-	
AA (Acceptance of assessment)	.201	.221	.348	.109	.322	.155	.185	-

All Pearson's r correlation coefficients are significant at $p < 0.01$, except for (n.s.)
n.s. not significant

5.3 Teachers' assessment profiles

Now that the latent factorial structure of the questionnaire was established, we could investigate whether teachers' views on assessment can be characterized by assigning the teachers to different assessment profiles. Therefore, we performed a latent class analysis on the item-level data. We estimated several models and in the end opted for the best fitting solution with three classes (cf. the lowest value of the BIC, Figure 1). The relative entropy of .917, which provides an indication for the uncertainty of the classification (where 1 is low uncertainty and 0 high), was near 1, indicating that the three latent classes were clearly separated.

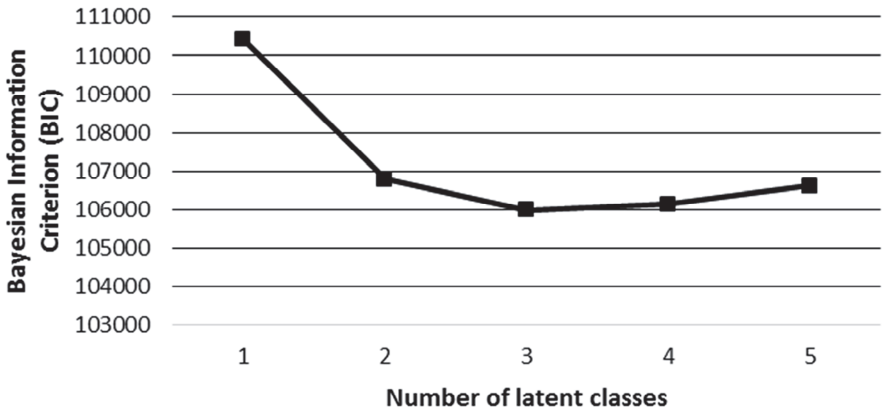


Figure 1. The value of the Bayesian Information Criterion (BIC) for one to five latent classes.

Based on this latent class analysis we then investigated whether teachers assigned to the three different latent classes differed on the eight factors that were identified in the questionnaire. The results clearly show that teachers from the different latent classes differed significantly from each other. We found large effects for *General instructional decision-making assessment purposes* ($F(2,1025) = 350.7$, $p < .001$, $\eta^2 = .406$) and *Specific instructional decision-making assessment purposes* ($F(2,1025) = 310.7$, $p < .001$, $\eta^2 = .377$). For *Assessment methods* ($F(2,1025) = 87.8$, $p < .001$, $\eta^2 = .146$), *Importance of assessing skills and knowledge* ($F(2,1025) = 194.7$, $p < .001$, $\eta^2 = .275$), *Importance of assessing extra-curricular skills* ($F(2,1025) = 85.0$, $p < .001$, $\eta^2 = .142$), and *Perceived usefulness of assessment* ($F(2,1025) = 171.1$, $p < .001$, $\eta^2 = .250$), the effects were small to medium in size. The effects were very small for *Diversity of assessment problem format* ($F(2,1025) = 12.7$, $p < .001$, $\eta^2 = .024$) and *Acceptance of assessment* ($F(2,1025) = 4.5$, $p = .011$, $\eta^2 = .009$). Post-hoc tests using Bonferroni correction showed that the differences between all three latent classes were significant for *General instructional decision-making assessment purposes*, *Assessment methods*, *Importance of assessing skills and knowledge*, *Importance of assessing extra-curricular skills*, and *Perceived usefulness of assessment* (all *ps* $< .001$), with the first latent class having higher scores on these factors than the second, and the second than the third (see also Figuer 2 for these comparisons). Having a higher score means, for example, that teachers used various assessment methods more often or hold a more positive view on assessment. The differences between the first and the second class were not significant for *Specific instructional decision-making assessment purposes* ($p = .082$) and *Diversity of assessment problem format* ($p = .076$), but these two classes did differ significantly from the third latent class ($ps < .001$). Finally, on *Acceptance of assessment* the second latent class scored significantly higher than the third ($p = .026$) but the other differences were not significant ($p = .091$ and $p = 1.00$). Figure 2 shows the profiles of teachers from the three different classes in relation to the eight standardized measures of teachers' views on mathematics assessment.

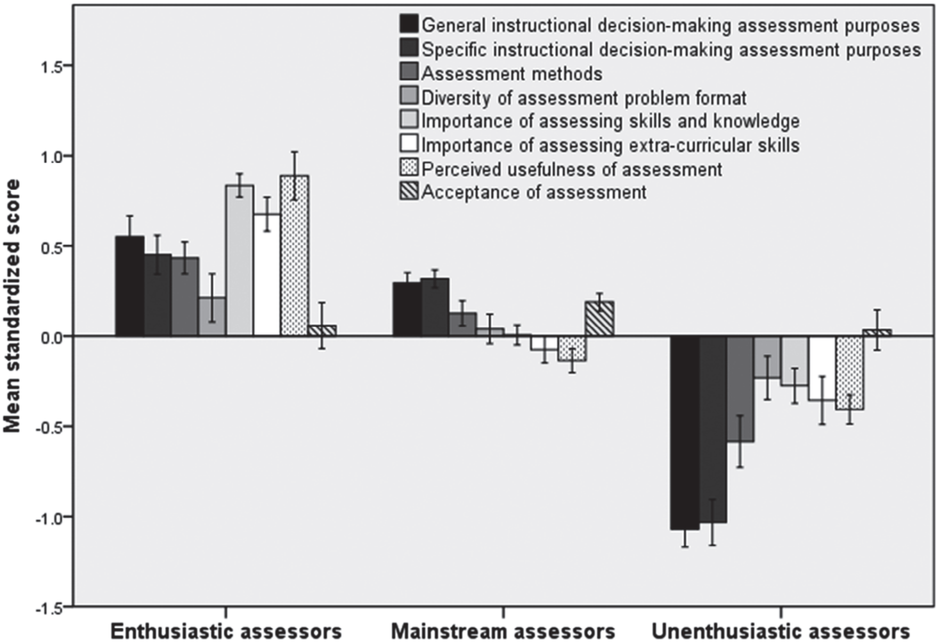


Figure 2. Mean standardized scores on factors for teachers in the three latent classes. Whiskers indicate 95% confidence interval.

We interpret the resulting assessment profiles as follows. The teachers belonging to the first class (21.7%) had above average scores on almost all factors. As these teachers reported to often use assessment for a variety of purposes, with frequently different assessment methods, acknowledged the importance of assessing skills and knowledge, and perceived assessment to be useful, we considered these teachers to be *enthusiastic assessors*. The biggest group of teachers (53.1%) formed the second class. These teachers scored quite close to the mean on all factors and relatively high on *Acceptance of assessment*, so we called them *mainstream assessors*. Teachers in the third class (25.2%) were considered *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 important or useful.

In Table 10, the standardized means per profile for the eight factors of the questionnaire and the means on background variables are displayed. We found that there were no significant differences between the teachers with different assessment profiles in terms of their age ($F(2,1069) = 1.00$, $p = .370$), the number of students in their classes ($F(2,1071) = 2.89$, $p = .057$), and whether they had at least a Bachelor's degree ($\chi^2(2, N = 1082) = 2.25$, $p = .324$). Mainstream assessors ($M = 13.8$, $SD = 8.4$; $F(2,1072) = 3.49$, $p = .031$) had significantly more teaching experience than Unenthusiastic assessors ($M = 12.2$, $SD = 8.5$; $p = .023$, $d = 0.190$). There was a significant relation between the teachers' gender and assessment profile ($\chi^2(2, N = 1089) = 15.7$, $p < .001$), with proportionally more female *Enthusiastic assessors* (91%) than *Mainstream assessors* (85%) and *Unenthusiastic assessors* (79%). With Kruskal-Wallis tests, we found that the frequency with which teachers discussed the learning goals with students was significantly related to their assessment profile ($\chi^2(2, N = 1076) = 70.5$, $p < .001$). *Enthusiastic assessors* discussed their learning goals more often than *Mainstream assessors* and these more frequently than *Unenthusiastic assessors*. These same significant differences were apparent between the assessment profiles in relation to the frequency with which teachers divide their students in level groups ($\chi^2(2, N = 1052) = 35.1$, $p < .001$) and the frequency with which they assess to get new information ($\chi^2(2, N = 1061) = 194.9$, $p < .001$).

Table 10

Mean values of factors and related variables for teachers in the three assessment profiles

Variables	Total	Assessment profiles			Sig. diff.
		1 Enthusiastic	2 Mainstream	3 Unenthusiastic	
General decision-making assessment purposes (z)		**0.55**	0.27	-1.07	1 > 2 > 3
Specific decision-making assessment purposes (z)		**0.42**	0.31	-1.03	1,2 > 3
Assessment method (z)		**0.41**	0.11	-0.61	1 > 2 > 3
Diversity assessment problem format (z)		**0.21**	0.02	-0.23	1,2 > 3
Importance of assessing types of skills and knowledge (z)		**0.80**	-0.11	-0.46	1 > 2 > 3
Importance of assessing extra-curricular skills (z)		**0.67**	-0.10	-0.38	1 > 2 > 3
Perceived usefulness of assessment (z)		**0.87**	-0.15	-0.45	1 > 2 > 3
Acceptance of assessment (z)		0.04	**0.10**	-0.23	2 > 1,3
Number of teachers	1101	239	585	277	
Age (years)	36.1	35.6	36.3	35.9	n.s.
Gender (% female)	85	**91**	85	79	1 > 2 > 3
Bachelor's or Master's degree (%)	30	34	28	29	n.s.
Teaching experience (years)	13.3	13.2	13.8	12.2	2 > 3.
Students in class (mean number)	53.3	52.2	54.5	51.9	n.s.
Frequency of discussing goals with students[a]	4.1	**4.5**	4.2	3.5	1 > 2 > 3
Frequency of assessing for information[a]	5.4	**6.1**	5.5	4.6	1 > 2 > 3
Frequency of making new level groups[a]	3.51	**3.80**	3.57	3.15	1,2 > 3

Note. The significantly highest value per row is printed in **bold**.
n.s. not significant
[a] Mean scores of response options: 1 = Rarely to never, 2 = Yearly, 3 = A few times a year, 4 = Monthly, 5 = Weekly, 6 = A few times a week, 7 = Daily.

6. Discussion

Three assessment profiles of Chinese primary school mathematics teachers were identified in this study. Teachers in these different profiles had distinct characteristics regarding their views on assessment. More than half of the teachers of our sample belonged to the profile of *Mainstream assessors*. These teachers appeared, as the name also indicates, to be moderate in their use of assessment. They reported to use several assessment methods for different purposes of instructional decision-making with an average frequency. To assess students, they reported using a number of different problem formats. These *Mainstream assessors* generally also 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 the use of assessment in their practice. The second group of teachers contained about one fifth of the sample and were *Enthusiastic assessors*. These teachers had above average scores overall. They reported to use 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. In addition, they reported to share learning goals with their students, to adjust the level groups in which the students are placed about monthly, and to collect information about student learning a few times per week, which was more often than the other two assessment profiles.

Taking the *Mainstream* and *Enthusiastic assessors* together, it shows that a large proportion of Chinese primary school mathematics teachers reported to use a variety of assessment methods for different purposes of supporting teaching and learning. These reported practices are in line with what is suggested in the Chinese mathematics curriculum standards about assessment (MoE, 2011). Also, teachers in many other countries have reported these practices (e.g. Krzywacki et al., 2011; Riggan & Oláh, 2011; Suurtamm et al., 2010; Veldhuis et al., 2013; Veldhuis & Van den Heuvel-Panhuizen, 2014b). Furthermore, the teachers with these two profiles generally agreed that assessment is useful for improving teaching and for enhancing learning, which was also found in several other countries (Brown, 2004, 2009; Brown et al., 2015; Brown, Lake et al., 2011).

Contrastingly, the teachers in the third profile reported remarkably different views on assessment, and were therefore called *Unenthusiastic assessors*. About one quarter of the teachers of our sample were in this profile, holding generally negative views on assessment. These teachers scored overall far below the mean, which reflects that they neither reported to use assessment purposefully or regularly, nor deemed it to be important or useful.

When looking at teachers' views on the influence of assessment on their teaching, nearly all reported to find assessment useful for their teaching, but nonetheless, one third reported that assessment did not influence their teaching. This finding is in line with what was uncovered in a review of Chinese teacher-written papers on assessment (Zhao et al., 2017, *Chapter 2 of this thesis*). In that review, it was found that, although the teachers made clear that one of the main purposes of doing assessment is improving teaching, they hardly reflected on adapting their further teaching based on assessment information.

When interpreting and using the results of this survey, a number of limitations need to be taken into account. Firstly, despite that the final sample included a considerable number of Chinese primary school mathematics teachers, compared to the large population in mainland China, we only had a relatively small sample. Another shortcoming is that we did not use a random sampling method but recruited teachers from the first author's circle of acquaintance, which may have increased the chance for getting not representative findings. However, through this method we ended up with –in an absolute sense– quite large sample of 1101 teachers which may have lowered the chance of getting biased findings. Yet, this does not mean that we think our sample is representative for all teachers in China. Although our sample covers teachers from various regions in China, it turned out that the teachers are mainly from Hebei, a province with an above average educational development level. It remains unsure whether teachers from other places have the same views on assessment, since the educational situation in China can be very different between regions. So, we should be prudent with connecting firm conclusions to our findings. Finally, because the findings are based on teachers' self-reported data, further direct sources like carrying out classroom observations could provide more insight into what really goes on in their classrooms. Besides, due to the fact that traditional external examinations in primary education are not officially used in some districts in Mainland China, no questions about this issue were included in the questionnaire. Yet, how primary school mathematics teachers' views on assessment are related to, or influenced by, the traditional external examinations, is still worthy to be explored.

In sum, despite the limitations of this survey, it provided us with relevant information about primary school mathematics teachers' assessment profiles in China. This sheds light on what these teachers think of assessment and how they perceive their assessment practice. Through the teacher assessment profiles we could gain a general picture about the presence of particular assessment cultures as reflected in the teachers' responses. This picture clearly showed that one quarter of the teachers did not report to use assessment purposefully or regularly, and did not deem it to be important or useful. A possible explanation for this negative view of assessment is that only 3% of the teachers reported that they had received professional development related to assessment. Notwithstanding this lack of professional development on assessment, the *Mainstream* and *Enthusiastic assessors* did have a positive approach to assessment and also reported to use it. Being able to identify these teachers and making use of their knowledge and experience can be a first step towards further development of an assessment practice that supports learning.

Acknowledgements

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

References

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge School of Education.

Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65*, 107–116.

Barrett, P. (2007). Structural equation modelling: Adjuging model fit. *Personality and Individual Differences, 42*(5), 815–824.

Berry, R. (2011). Educational assessment in mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 49–61). Netherlands: Springer.

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

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301–318.

Brown, G. T. L. (2009). Teachers' self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teo & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications* (pp. 243–266). Rotterdam, NL: Sense Publishers.

Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' selfreported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research, 71*, 50–64.

Brown, G. T. L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education, 2*(1), 993836.

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.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347–363.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*(1), 210–220.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(1), 1–6.

Chen, J., & Brown, G. T. L. (2016). Tensions between knowledge transmission and student-focused teaching approaches to assessment purposes: Helping students improve through transmission. *Teachers and Teaching: Theory and Practice, 22*(3), 350–367.

Dias, J. G., & Vermunt, J. K. (2006). Bootstrap methods for measuring classification uncertainty in latent class analysis. In A. Rizzi & M. Vichi (Eds.), *Proceedings in computational statistics* (pp. 31–41). Heidelberg, Germany: Springer.

Harlen, W. (2007). *Assessment of learning*. London, UK: Sage.

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.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

IBM Corp. (2014). *IBM SPSS statistics for windows, version 23.0*. Armonk, NY: IBM Corp.

Krzywacki, H., Koistinen, L., & Lavonen, J. (2011). Assessment in Finnish mathematics education: Various ways, various needs. In B. Ubuz (Ed.), *Proceedings of the 35th annual conference of the international group on the psychology of mathematics education* (1st ed., p. 340). Ankara, Turkey: International Group on the Psychology of Mathematics Education.

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*, 12–23.

Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process — the assessment episode. *Assessment in Education: Principles, Policy, & Practice, 4*(3), 381–399.

Ministry of Education of the People's Republic of China (MoE). (2001). *Quanrizhi yiwu jiaoyu shuxue kecheng biaozhun (shiyangao)* [Mathematics curriculum standards of nine-year compulsory education (trial version)]. Beijing, China: Beijing Normal University Press.

Ministry of Education of the People's Republic of China (MoE). (2011). *Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

Ministry of Education of the People's Republic of China (MoE). (2014). *2013 nian jiaoyu tongji shuju* [Educational statistics data in 2013]. Retrieved from http://www.moe.edu.cn/s78/A03/moe_560/s8492/s8493/201412/t20141218_181947.html

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 369–377.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide (Sixth ed.)*. Los Angeles, CA: Muthén & Muthén.

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.

Nortvedt, G. A., Santos, L., & Pinto, J. (2016). Assessment for learning in Norway and Portugal: The case of primary school mathematics teaching. *Assessment in Education: Principles, Policy & Practice, 23*(3), 377–395.

OECD (2012). *Education at a Glance 2012: OECD Indicators*. Retrieved from http://www.oecd.org/edu/EAG%202012_e-book_EN_200912.pdf .

Pan, Z. (2015). Xiaoxue jieduan "tongkao tongce" de zaisikao [Rethinking of standardized tests in primary education]. *Jiaoyu celiang yu pingjia (Lilunban), 7*, 63–64.

Phelan, J., Choi, K., Vendlinski, T., Baker, E., & Herman, J. (2011). Differential improvement in student understanding of mathematical principles following formative assessment intervention. *The Journal of Educational Research, 104*(5), 330–339.

Riggan, M., & Oláh, L. N. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment, 16*(1), 1–14.

Sach, E. (2012). Teachers and testing: An investigation into teachers' perceptions of formative assessment. *Educational Studies, 38*(3), 261–276.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758–765.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271–286.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399–417.

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014a). 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. (2014b). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE, 9*(1), e86817.

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.

Wang, S., Yuan, L., Tian, Z., & Zhang, X. (2013).Woguo geshengfen jiaoyu fazhan shuiping bijiao fenxi [The comparative research on the level of regional development of education]. *Jiaoyu yanjiu, 6*, 29–41.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65.

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.

Authors' contributions

This paper was a collaborative work of the three authors. XZ, MHP, and MV all participated in designing the study. MHP and MV developed the instrument. XZ adapted the instrument, carried out the data collection in China, and was responsible for interpreting the Chinese resources. XZ, MHP, and MV analyzed the data, and drafted and revised the manuscript. All authors read and approved the final manuscript.

Chapter 4

Teachers' use of classroom assessment techniques in primary mathematics education — an explorative study with six Chinese teachers

Xiaoyan Zhao[a], Marja van den Heuvel-Panhuizen[a,b], & Michiel Veldhuis[b]

[a]*Freudenthal Institute, Faculty of Science, Utrecht University, the Netherlands*

[b]*Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*

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), 1-18.

Teachers' use of classroom assessment techniques in primary mathematics education — an explorative study with six Chinese teachers

Abstract

This paper reports on the use of classroom assessment techniques (CATs) by primary school mathematics teachers in China. CATs are short, focused assessment activities that can reveal students' understanding of specific mathematical subjects. The study involved six female third-grade mathematics teachers from Nanjing, China. The focus was on assessing division. Data were collected by teacher interviews, feedback forms and final reports, lesson observations, and student work. The study revealed that the teachers could easily include CATs in their daily practice. By conducting the CATs, the teachers got new information about their students' learning. Most teachers liked using the CATs, especially those with the red/green cards, which is a whole-classroom immediate response format, providing quick information of the students' learning. The teachers also found the CATs feasible to conduct and helpful to engage their students during the lesson. However, no evidence was found that they used the information gained from the CATs for adapting their instruction to meet the students' needs in subsequent lessons. In fact, the teachers only used the teacher guide of the CATs to adapt their instruction beforehand. The CATs, instead of being implemented as assessment activities, were often included as extra exercises in the pre-arranged lesson plans of the teachers. If necessary, the teachers provided their students with instant help in order to assist them to get the correct answers. In general, the teachers were positive about the CATs as a way to reveal their students' understanding of division in an effective and efficient fashion. The teachers recognized that it can be very revealing to challenge their students with questions that are not completely prepared by the content of their textbooks. The results of this study suggest that on the one hand CATs can be helpful for Chinese mathematics teachers' formative assessment practice in primary education. On the other hand, our study also provides some evidence that using CATs, as an approach to formative assessment, to make informed and adequate decisions about further teaching, can be a real challenge for teachers.

Keywords: Classroom assessment; mathematics education; primary school; teachers; China

1. Background

Classroom assessment, as formative assessment in the hands of teachers with the aim of collecting information about the students' learning to make adequate instructional decisions to meet the students' needs, has been widely acknowledged and promoted in the field of education. In mathematics education in China, the idea of using assessment to support teaching and learning has also become the centerpiece of the assessment reform since 2001. However, after over 10 years of effort, studies showed that primary mathematics teachers still have difficulties in implementing assessment in their classroom practice. The current study was set up to explore whether classroom assessment techniques (CATs), which are short and focused assessment activities carried out by the teacher for revealing students' understanding of specific mathematical topics, have potential in the context of Chinese primary mathematics education.

1.1 Classroom assessment

Knowledge of what students know is indispensable for educational decision-making. This is true at all levels of education, from kindergarten to university, and from the micro-setting of a classroom to the macro-environment of educational policy. Without information about student learning, the educational system cannot function. Therefore, assessment, as the process in which students' responses to specially created or spontaneously occurring stimuli are collected to draw inferences about the students' knowledge and skills (Popham, 2000), plays a key role in education. Depending on the purpose assessment is used for, in education, two main types of assessment are distinguished: formative assessment and summative assessment. Formative assessment is an interim-assessment to find clues for further instruction. Therefore, formative assessment is considered as assessment *for* learning and is often contrasted with assessment *of* learning (e.g., Wiliam, 2011a), which refers to summative assessment that aims to evaluate a student's learning at the end of an instructional sequence to give the student a mark or a certificate.

Although, according to some authors (e.g., Black, 2013; Harlen, 2005), formative and summative assessments should not be seen as separated entities or different types of assessment, because they are both important for evoking information about knowledge, understanding, and attitudes of students, in this paper we only focus on formative assessment. We consider formative assessment as the assessment that teachers continuously do during teaching: figuring out what their students know or what difficulties their students have, and using this knowledge to adapt their instruction to cater for the students' needs. This assessment in the hands of teachers with the aim to make decisions about the next step in instruction is often called classroom assessment (e.g., Shepard, 2000). In this, it is recognized that the teachers, rather than particular outsiders, are in the best position for eliciting and collecting adequate and quality information about their students' learning (Harlen, 2007). Classroom assessment can only function formatively when the collected information is actually used by the teacher to adapt the teaching to meet students' needs (Black & Wiliam, 1998a). With respect to the actions taken by the teachers, a distinction can be made between enhancing students' performance by correcting students' responses immediately and instantly explaining why the answer is wrong, or by a postponed action by tailoring their instruction to the needs of the students and in this way improving the students' learning (Antoniou & James, 2014; Hill & McNamara, 2012).

1.2 Classroom assessment techniques

Since Black and Wiliam (1998a, 1998b) brought the power of classroom assessment to raise students' achievement to a larger audience, more research has been conducted on its practical applications. Leahy, Lyon, Thompson, and Wiliam (2005) provided teachers with various activities to improve their classroom assessment practice. Based on the teachers' tryouts, these researchers came to more than 50 assessment techniques. Typical for these techniques is that they blur the divide between instruction and assessment, and make it possible to adjust the teaching while the learning is still taking place. Another characteristic of these techniques is that they are low-tech, low-cost, and often well-known activities done by teachers, which require only subtle changes in practice and can be feasibly implemented by teachers. For example, in daily teaching, to make the decision whether to go over something once more or to move on, teachers need to have insight into students' thinking. Wiliam (2011b) proposed to use range-finding questions and hinge-point questions to assess what students already know at the beginning of, or during, the lesson. Moreover, in order to avoid deciding for the whole class based on the performance of just a few students, ABCD cards, through which individual students can show their answers by raising a card, and exit passes, which means that students have to solve some problems in a worksheet before leaving the classroom, were recommended.

In the work done by Wiliam and his colleagues (Leahy et al., 2005; Wiliam, 2011b), they shared the techniques with teachers who taught different subjects from different educational levels and found that the techniques were useful in supporting teachers' effective formative assessment across content areas and age brackets. This finding of Wiliam and his colleagues is encouraging. However, it is also natural and reasonable to consider that those techniques must be content-dependent. After all, what is really asked by teachers as a range-finding or hinge-point question and the problems in the exit pass worksheets matters most for what information about students' learning can be elicited.

Inspired by the work of Wiliam and his colleagues, also in the Netherlands, studies were set up to investigate the use by primary school teachers in mathematics of what were called classroom assessment techniques (CATs) (Veldhuis & Van den Heuvel-Panhuizen, 2014, 2017). These CATs, similar to the ones William and his colleagues used (see Leahy et al., 2005; Wiliam, 2011b), were short and focused assessment activities carried out by the teacher with the purpose of revealing students' understanding of specific mathematical subjects. In using these CATs, teachers could collect information about their students' learning, thus allowing them to adapt their subsequent teaching to meet their students' needs. To develop the CATs, first a textbook analysis was performed, since assessment should be closely connected to the mathematics currently taught in class in order to make classroom assessment information useful for teachers. However, this connectedness to the textbook does not mean that the CATs merely repeated the tasks that are in the textbook. Instead, the CATs provide students with new questions or tasks that can reveal their deep knowledge of a particular concept from a different perspective. In addition to the content, also decisions have to be made regarding the format of the CATs to make sure students' learning information can be assessed by the teacher in an efficient and effective way. Two main formats were employed. By using the format of the red/green cards, in which students show their answers by holding up a colored card, teachers can easily discover the students that have the correct answer and those who do not. Additionally, the way in which students raise their card – whether they react quickly and with confidence or they hesitate to respond or change their card after they have seen others' cards – is also valuable information. When it was more desirable to have detailed information about the students' thinking steps based on their written responses, the format of worksheets was employed.

The studies conducted in Dutch primary schools were meant to qualitatively investigate the feasibility of the CATs and experimentally evaluate the effectiveness of the teachers' use of the CATs. In two pilot studies, ten primary school teachers and over 200 students in Grade 3 were involved (Veldhuis & Van den Heuvel-Panhuizen, 2014). Although the teachers were offered a collection of CATs, they were free in changing these CATs or making their own CATs in order to have them fit their classroom situation. Results from these pilot studies showed that teachers and students enjoyed using the CATs and found them useful. Moreover, the students whose teachers used the CATs improved considerably more in their mathematics achievement as measured by a standardized mathematics test than the students in a national reference sample did. Later on, the effectiveness of teachers' use of the CATs on students' achievement was further confirmed in a large-scale quasi-experimental study with 30 primary teachers and 616 students (Veldhuis & Van den Heuvel-Panhuizen, 2017).

1.3 Classroom assessment in China

In 2001, in China, which has a long history of examination oriented education (Berry, 2011), a new approach to assessment was introduced as part of the New Curriculum Reform that was launched by the Ministry of Education (MoE, 2001). To reduce the overemphasis on grading and ranking – which was common practice before the reform – it was emphasized in the mathematics curriculum standards that

> [t]he main purpose of assessment is getting the whole picture of process and outcomes of students' mathematics learning, stimulating students to learn and improving teachers' instruction (MoE 2011, p. 52).

This means that instead of using only externally developed standardized tests for assessing students, teachers are now the key stakeholders in implementing assessment policies (Yu & Jin, 2014). To better support teachers to perceive and practice this new idea of assessment, also the mathematics curriculum standards document provide guidelines, namely about the content of assessment, the person who can be the assessor, the methods that can be used for assessment, and the ways of reporting and using assessment results (MoE, 2011). It is stipulated that assessment should address what mathematics students have to learn and what mathematical competences they have to develop, regarding their knowledge and skills, mathematical thinking and problem solving, and mathematical and learning attitude. For example, the assessment of mathematical thinking and problem solving should be carried out by multiple methods during the whole process of mathematics learning. Although teachers are undoubtedly playing an important role in assessment, students and their peers are also encouraged to be actively involved in the assessment activities. In the assessment guidelines, assessment methods like oral tests, open questions, observations, exercises in and after class, and many more are suggested to be used in the assessment of students' learning. Finally, in terms of reporting assessment results, teachers are recommended to provide students with feedback that focuses on what the students learned, the progress they made, their potential, and where they need to improve. Based on the information about the students' learning level and their learning difficulties, teachers are suggested to adapt and improve their instruction.

Since 2001, great effort has been made to put assessment into teachers' hands by helping them to employ the new idea of assessment and enhance their assessment ability, as stated by Zhang (2009). However, after a decade, it was found in a large-scale questionnaire survey study (Brown, Hui, Yu, & Kennedy, 2011), in which 898 teachers from Southern China were involved, that teachers seemingly held the view that such assessment was only weakly relevant to real improvement in teaching and learning. Moreover, some researchers (Cui, 2008; Zhong, 2012) pointed out that Chinese teachers are still used to pay much more attention to what and how they teach than to what and how they assess. Recently, Zhao, Van den Heuvel-Panhuizen, and Veldhuis (2017a, *Chapter 2 of this thesis*) conducted a literature review based on 266 papers on classroom assessment written by Chinese primary mathematics teachers. In this review, it was found that the teachers overlooked using assessment information to adapt and improve their further instruction. Furthermore, in a large-scale questionnaire survey (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2017b, *Chapter 3 of this thesis*) on teachers' assessment practice and beliefs in primary mathematics classes in China, it was revealed, based on 1101 [1] teachers' responses, that teachers did not consider questioning as relevant enough to provide useful student learning information, despite assessing their students by questioning nearly every day.

1.4 Possible usefulness of CATs in China

Although the aforementioned studies, of course, cannot be considered as providing a full picture of the classroom assessment practice of primary school practice in China, they offer at least some evidence that the teachers' assessment practice can be improved and that it can be brought more in agreement with the assessment as suggested in the curriculum standards (MoE, 2011). A possible way might be the use of CATs. In the first place, because the conceptualization of CATs is quite in line with the approach to assessment that is advocated in the Chinese assessment guidelines. The use of CATs could provide Chinese teachers with clear and concrete examples of how to employ questioning to dig out students' mathematical understanding. Moreover, the formats of CATs, especially by using red/green cards, may invite more students to actively participate in assessment activities. Also, it is worthwhile to note that CATs can be used in a whole-class setting to collect information quickly and easily from a large group of students, a feature that corresponds quite well to the average Chinese classroom situation with about 37 students in one class (OECD, 2012, p. 450). According to Zhao, Mulligan, and Mitchelmore (2006), a large class size is one of the principal reasons for the gap between the actual assessment practice and the intended assessment in official curriculum documents. A further reason for introducing CATs to Chinese primary school teachers is that several studies in other countries (Leahy et al., 2005; Veldhuis & Van den Heuvel-Panhuizen, 2014, 2017; Wiliam, 2011b) have shown that these focused and short assessment activities, initiated by the teacher and aimed at revealing students' understanding of a particular aspect of mathematics, were helpful for teachers to assess their students. However, positive experiences with CATs in one country do not necessarily imply that they are also feasible and effective in other countries. What would be a good approach to formative assessment may be different in countries with different approaches to teaching

[1] In the published paper, the number of teachers' responses is 1158, but it should have been 1101 (see *Chapter 3 of the thesis*).

and different classroom practices (e.g., Shepard, 2000). Studies have revealed that culture matters in mathematics education and that there are differences between mathematics education in, for example, East Asian countries and Western countries (Leung, Graf, & Lopez-Real, 2006). So we are not sure whether CATs are useful for Chinese primary school mathematics teachers. Therefore, the current study intended to disclose what the potential of this approach to formative assessment for the Chinese context could be. More in particular, the research questions of our study were:

1. How are the CATs used in the context of Chinese primary mathematics education?

2. What information do the teachers who use the CATs get from CATs and what do they do with this information?

3. Do these teachers think CATs are useful and do they want to use CATs in the future?

2. Method

To answer the research questions, an explorative study, applying a case study approach, was carried out in which Chinese primary school mathematics teachers put into practice a package of CATs. The CATs were attuned to the mathematics textbook that the teachers in Grade 3 used to plan their teaching. The teachers worked with the package in February – March 2014, which was the beginning of the second semester.

2.1 Participants

The study was carried out in Nanjing, which was the city where the first author studied and knew a number of schools. Five schools were contacted, and two of them were willing to participate. These schools are located in the urban area of Nanjing. All third-grade mathematics teachers in these two schools agreed to be involved and chose one of their two classes to take part in the study. The convenience sample we got in this way consisted of six female teachers with the average teaching experience of over 9 years (minimum 1 year; maximum 25 years) and their 216 students. Teachers A and B were from School I, which is a school with an average reputation, and in their classes there were around 30 students. Teachers C, D, E, and F came from School II, which has a good reputation for its quality of education and has better facilities than School I, and they had about 39 students in their classes. All six teachers involved in this study are specialist teachers, they only teach mathematics. They had been teaching their students for at least one semester, which means they were all familiar with their students' learning situation.

2.2 Used textbook

The textbook was the main reference for designing the CATs, because Chinese mathematics teachers rely heavily on textbooks as the main resource for their day to day instruction (Li, Chen, & Kulm, 2009) and pay much attention to study and understand textbooks carefully and thoroughly (Cai & Wang, 2010). In this way, the CATs could be embedded in the teachers' daily classroom practice. The six teachers all used the Sujiaoban (苏教版) textbook, published by Jiangsu Education Publishing House (2005).

Based on the characteristics of CATs developed in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014), new CATs were designed that fitted the content and teaching of the Chinese textbook. At the beginning of the second semester of Grade 3, one of the addressed content domains in this textbook is division. The focus in this paper is on this domain.

2.2.1 Teaching trajectory for division

To illustrate how the teaching of division is built up and how division is connected to the related mathematical domain of multiplication, Table 1 shows the teaching trajectory for multiplication and division in the Sujiaoban textbook. Although the study focused on Grade 3, to provide a long-term overview, the table shows the trajectory from Grade 2 to Grade 4.

Table 1

Multiplication and division in 苏教版 textbook (Grade 2 – Grade 4)

Grade-Semester	Chapter	Topic	Content	Example problem: Horizontal notation		Example problem: Vertical notation	
2-1	1	Multiplication	Meaning of multiplication; symbol (×)	4×2			
	2	Multiplication	Multiplication tables from 1 to 6 (multiplier ≤6 and multiplicand ≤9)	1×1	6×9		
	4	Division	Meaning of division; symbol (÷)	6÷2			
	5	Division	Division problems with divisors from 1 to 6 (dividend ≤9 times divisor)	10÷2	36÷6		
	8	Multiplication and division	Multiplication tables from 7 to 9 (multiplier and multiplicand ≤9) and related division problems	6×7 9×9	42÷7 81÷9		
			Vertical notation of multiplication and division			$\begin{array}{r} 4 \\ \times\ 2 \\ \hline 8 \end{array}$	$\begin{array}{r} 3 \\ 2\overline{)6} \\ 6 \\ \hline 0 \end{array}$
2-2	1	Division	Division with remainder; 1- or 2-digit divided by 1-digit	7÷3	17÷5	$\begin{array}{r} 2 \\ 3\overline{)7} \\ 6 \\ \hline 1 \end{array}$	$\begin{array}{r} 3 \\ 5\overline{)17} \\ 15 \\ \hline 2 \end{array}$
	8	Multiplication	2-digit multiplied by 1-digit	20×3			
				14×2	48×2	$\begin{array}{r} 14 \\ \times\ 2 \\ \hline 28 \end{array}$	$\begin{array}{r} 48 \\ \times\ 2 \\ \hline 96 \end{array}$

Grade-Semester	Chapter	Topic	Content	Example problem: Horizontal notation	Example problem: Vertical notation	
3-1	1	Division	2-digit divided by 1-digit	40÷2		
				52÷2 62÷3	26 2)52 4 12 12 0	20 3)62 6 2
	7	Multiplication	3-digit multiplied by 1-digit	400×2		
				120×4	152 × 4 608	120 × 4 480
3-2	1	Division	3-digit divided by 1-digit	600÷3		
				306÷3	78 4)312 28 32 32 0	102 3)306 3 6 6 0
	4	Multiplication	2-digit multiplied by 2-digit	12×10		
				25×30	28 × 12 56 28 336	25 × 30 750
4-1	1	Division	2- or 3-digit divided by 2-digit	60÷20	3 20)60 60 0	8 34)272 272 0
4-2	1	Multiplication	3-digit multiplied by 2-digit		144 × 15 720 144 2160	850 × 20 17000

The teaching of multiplication and division starts in the beginning of the first semester of Grade 2. The meaning of multiplication is introduced as repeated addition (Chapter 1) and that of division as equal sharing and equal grouping (Chapter 4). A group of objects is the main model that is used to support students in their understanding of the meaning of multiplication and division. Later, the multiplication tables and related division problems become the focus of learning, and ratio tables appear as an important tool (Chapters 2, 5, and 8). At the end of the first semester of Grade 2 (Chapter 8), the algorithmic approach for solving multiplication and division problems is introduced. In Grades 3 and 4, solving multiplication and division problems with the algorithm becomes one of the main objectives. Students are expected to solve these problems with numbers with an increasing number of digits (see the Content column in Table 1).

2.2.2 Division of three-digit numbers by a one-digit number

The CATs developed for this study were based on the content of Chapter 1 that is taught in the second semester of Grade 3 (see the framed section in Table 1). Like the other chapters in the textbook, this chapter is organized around a series of example problems, which are introduced from contexts. Students first solve simple division problems by mental calculation. Then the focus turns to solving problems by using the algorithm. The main objective of the chapter is that students become able to solve division problems with a three-digit number divided by a one-digit number. Chapter 1 contains eight lessons in total (Table 2), including new lessons and review lessons. For each new lesson, the teacher guide of the textbook gives clear and specific objectives. For example, the objective for lesson 2 is that students need to be able to solve problems of three-digit numbers divided by a one-digit number resulting in a two-digit quotient and sometimes a remainder (decimal numbers are not yet introduced). In the textbook, the example problem of $312 \div 4$ is introduced within the context of selling eggs (Figure 1; the original text is in Chinese and is translated by the first author), together with exercises on bare number problems and context problems. In review lessons, students have to finish exercises of earlier lessons and do more comprehensive exercises. In lesson 8, which is a review lesson at the end of the chapter, exercises of mental calculation, calculation by using the algorithm, and more context problems are provided (Figure 2).

Table 2

Lessons in Chapter 1: division of three-digit numbers by a one-digit number

Lesson	Type of lesson	Type of division problem	Example problem	
			Horizontal notation	Vertical notation
1	New	3-digit divided by 1-digit with 3-digit quotient	600÷3=200	493 2) 986 8 18 18 6 6 0
2	New	3-digit divided by 1-digit with 2-digit quotient		78 4) 312 28 32 32 0
3	Review (Lesson 1-2)			
4	New	‘0’ in dividend (and quotient)	0÷3=0 306÷3=102	102 3) 306 3 6 6 0
5	New	‘0’ only in quotient		108 4) 432 4 32 32 0
6	New	Two-step division problem	There are two bookshelves with four layers. Suppose there are 224 books in total. How many books are there on one layer?	
7	Review (Lesson 4-6)			
8	Review (whole chapter)			

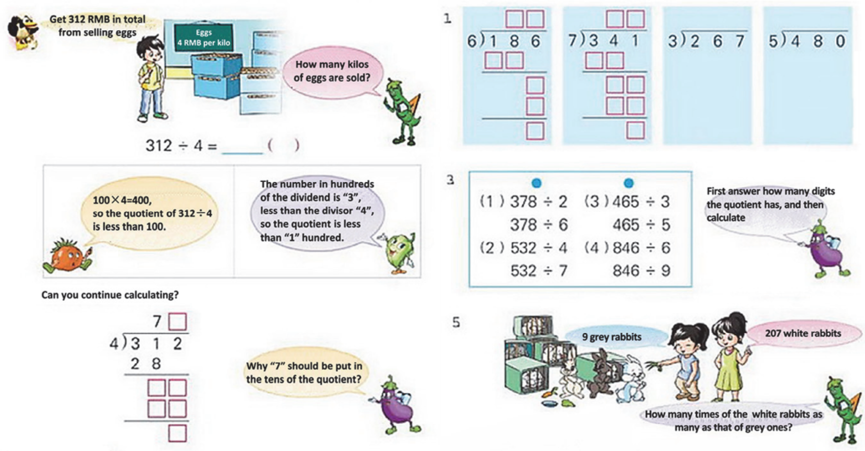


Figure 1. Example problem in Lesson 2 (left) and some corresponding exercises (right) (translated from Chinese)

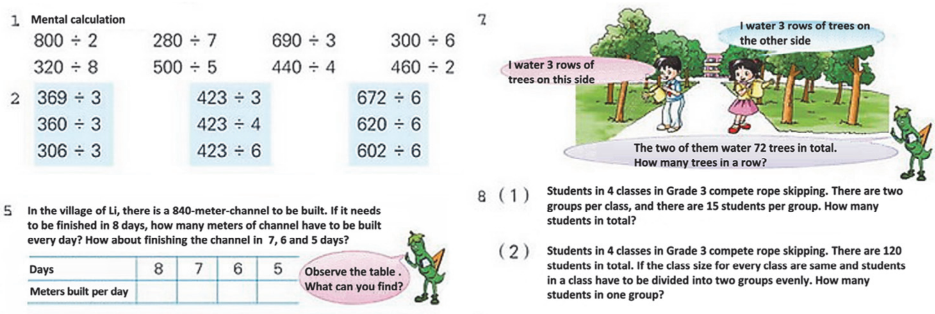


Figure 2. Part of the exercises in Lesson 8 (translated from Chinese)

2.3 CATs for division of three-digit numbers by a one-digit number

When designing the CATs, two requirements were taken into consideration. The CATs had to be linked to the objectives of the lessons included in the chapter. Moreover, the CATs had to provide teachers with information about their students' learning; in particular, the CATs should disclose information that could be useful for making decisions about further teaching. This means that in the CATs questions had to be asked that went beyond the regular textbook exercises and could reveal a deep level of understanding of division. In total, 13 CATs (Table 3) were developed for Chapter 1. Out of the 13 CATs, two are discussed in the following sections. The first one is a CAT with a whole-classroom immediate response format; the second one has an individual worksheet format.

Table 3
Descriptions of CATs for division of three-digit numbers by a one-digit number

Title	Format	Purpose
Connecting division to multiplication	Red/Green cards	Assessing whether students can find the related multiplication problem for a division problem
Family problems	Red/Green cards	Assessing whether students can recognize analogous problems and are aware of the relationship among the results of these problems
Choosing an answer for a division problem	Red/Green cards	Assessing whether students can estimate the quotient
Identifying the watershed	Red/Green cards	Assessing whether students can recognize the breaking point when the number of digits of the quotient changes
Checking divisibility	Red/Green cards	Assessing whether students have a clue in advance whether divisions have a remainder or not
Is it in the hundreds/tens?	Red/Green cards	Assessing whether students can estimate the quotient
Algorithm with ink blots	Worksheet	Assessing whether students understand how the division algorithm works
Is there a zero in the middle of the quotient?	Red/Green cards	Assessing whether students understand the relationship between the existence of zero in the dividend and in the quotient
Correct or not correct?	Red/Green cards	Assessing whether students can quickly check the correctness of the result of division problems without performing the algorithm
Is it in the multiplication table of …?	Red/Green cards	Assessing whether students have the multiplication number facts available
Possible remainders	Red/Green cards	Assessing whether students understand the relationship between divisors and remainders
Easy or difficult?	Worksheet	Assessing what is the easiness-difficulty range of students and whether they are aware what characteristics of a problem make it easy or difficult for them
Solving division problems without using the algorithm	Worksheet	Assessing whether students have a deep understanding of the division operation and whether they have, instead of the algorithm, other strategies available to solve division problems

2.3.1 Identifying the watershed (CAT-1)

CAT-1 was planned for lesson 2 in Chapter 1. In the previous lesson, the students were taught to solve problems in which a three-digit number is divided by a one-digit number with a three-digit quotient. Problems like 600 ÷ 3 have to be solved by using a horizontal notation, whereas for problems like 986 ÷ 2 a vertical notation is used (Table 2). In lesson 2, students also have to solve problems in which a three-digit number is divided by a one-digit number, but now the problems have quotients of two digits, like in 312 ÷ 4. The students have to solve these problems by carrying out the standard division algorithm using the vertical notation. An extra assignment given in the textbook for this lesson is that the students have to determine the number of digits in the quotient. The students have to find this number before they do the calculation (Figure 1, exercise 3). Normally, teachers call on individual students to give their answers. In this way, each division problem is dealt with separately and, consequently, this approach does not provide teachers with information about whether students know the underlying structure that determines the number of digits in the quotient and whether they have a more general understanding of the role of place value. CAT-1 (Figure 3) is meant to dig deeper into students’ understanding of the division operation. For this CAT, the red/green cards are used, which has a wholeclassroom immediate response format. In addition to Tasks 1 and 2, this CAT included two more tasks (Task 3: dividend 721; Task 4: dividend 7214). Teachers could vary the content of these tasks and the number of tasks they use.

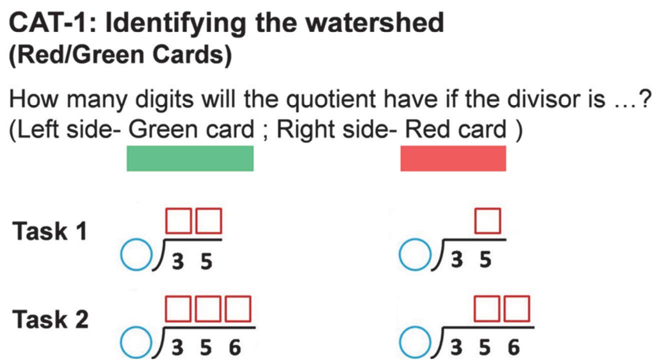


Figure 3. CAT-1: Identifying the watershed (translated from Chinese)

The teacher shows a division problem with the divisor left blank to the students, then mentions a series of possible divisors (increasing from 1 to 9). The students have to identify the breaking point when the number of digits in the quotient changes (that is, the watershed, because this change in the number of digits and consequently the color of the card, from green to red, is just like the divide in the flow of water that watershed refers to). In task 1, the dividend of the division problem is the two-digit number 35. On the left side, a problem is shown with a two-digit quotient, whereas, on the right side, there is a problem with a one-digit quotient. Both are possible. The students have to decide which card to raise when the teacher says: “35 divided by 1.” The green card stands for the quotient with two digits, and the red card represents the quotient with one digit. Then, the teacher moves on to the subsequent numbers as divisors (2, 3, 4, …). As the divisors get bigger and bigger, students can notice that from a particular divisor on (depending on the dividend), the number of digits of the quotient changes (the watershed point); till then students have to show the green card again and again and after reaching this particular divisor they can show the red card continuously. As a matter of fact, after passing the watershed point, no thinking is necessary anymore. The way students raise the cards may give teachers a quick first clue about whether students comprehend what determines the number of digits in the quotient.

2.3.2 Solving division problems without algorithm (CAT-2)

CAT-2 was planned for lesson 8 in Chapter 1. This is near the end of the chapter when most students are quite able to carry out the division algorithm and can solve the division problems presented in the textbook without mistakes. In CAT-2 (Figure 4), the students are asked to solve a number of division problems without using the standard algorithm. At first sight, this CAT looks like a contradiction in terms: assessing students understanding of division without asking them to perform the algorithm they have learned. However, the main idea of this CAT is that when students cannot solve a division problem without using the algorithm they will probably not have a good understanding of what a division really means. Even if students are able to perform every step of the algorithm without mistakes and arrive at the correct answer, this does not necessarily mean that they have a deep and stable understanding of the division operation. It is also possible that they just apply the procedure in a mindless, mechanistic way, which means that they might run into trouble when they to do more complicated division problems with, for example, decimal numbers. If, however, students do have this deep understanding of division, then they will also be able to use different strategies to deal with division problems, for example, by regrouping, using partitive and quotitative models, or thinking of the relation between multiplication and division. This is not to say that understanding the standard algorithm does not demand conceptual understanding of the division operation, for it does, or that the standard algorithm is not a worthwhile strategy, for it is, but merely using it does not necessarily imply deep understanding of division.

CAT-2: Solving division problems without algorithm
(Worksheet + later check by the teacher)

Here are some division problems.
Try to solve them without using the standard algorithm.
Write down how you proceed.

	How do you solve the problem?
Task 1) 468÷2	
Task 2) 594÷6	
Task 3) 480÷3	
Task 4) 816÷4	

Figure 4. CAT-2: Solving division problems without algorithm (translated from Chinese)

The format of CAT-2 is a worksheet. The teacher has to check student work after class, and then uses this information in the next lesson. The worksheet contains a small number of division problems presented as horizontal number sentences. Students are free in the way they solve the problems but are explicitly told not to use the division algorithm. Students who have a good understanding of division will be able to consider, for example, division as equal sharing, making groups, thinking about the relationship between multiplication and division, and can use this knowledge to solve the division problems without applying the algorithm.

2.4 Teacher support

To inform the teachers about the CATs, a teacher guide was developed describing for each CAT its purpose, how and when it can be used in class, and issues on which teachers can focus when observing and checking students' responses. The teacher guide of the CATs also provided some general background information about formative assessment and the characteristics of CATs. Although detailed instructions were given for using the CATs, the teacher guide of the CATs was not meant as a fixed recipe for what to do in class. Instead the teachers could adapt the CATs to their own needs, which is in line with the finding of Lee and Wiliam (2005) that having teachers decide for themselves about the use of assessment techniques is crucial for the success of their use. Thus, to enhance the implementation of the CATs and stimulate ownership, the teachers could freely decide which CATs to use, when, and in what way.

To further brief the teachers about the study, four 1-hour meetings were organized. They were led by the first author. The initial meeting took place 2 days before the teachers started with Chapter 1 and addressed the CATs used in the first week. In the next two meetings, new CATs and teachers' experiences with the previously used CATs were discussed. The last meeting was only dedicated to the teachers' reflection on using the CATs. A distinctive characteristic of the meetings was that the teachers helped each other to comprehend the essential aspects of the CATs and discussed how they might use them in their classroom. Sharing opinions of how to teach and deliberating their teaching plans collectively within schools is a rather common practice for teachers in China (Chen, 2006; Li & Zhao, 2011). This came also evidently to the fore in the meetings.

During the process of introducing CAT-1 and CAT-2, the content of the teacher guide of the CATs was explained. The teachers were encouraged to ask questions regarding these CATs. For example, some teachers wondered why, in CAT-1, task 4 (which has a four-digit dividend) was included, since it exceeded the learning scope of Chapter 1, in which students are only required to solve division problems up to three-digit dividends (Table 2). Nevertheless, the teachers thought CAT-1 was not difficult for most of their students. With respect to CAT-2, the teachers asked why students are asked to solve division problems without using the algorithm. The explanation given to them was that by offering students problems that differ from the exercises they normally do, teachers could get information about students' deep understanding of the division operation. Despite the fact that most of the CATs were new to the teachers, all the teachers expressed that they were willing to use them.

2.5 Data collection and data processing

The main method for the data collection was conducting teacher interviews. All teachers were interviewed at least two times by the first author. These interviews took place after the teachers gave a lesson in which they used a CAT. If a teacher was not interviewed, then she was asked to fill in a feedback form after the lesson with the CAT. The questions the teachers answered on this feedback form were the same as those asked in the interviews. At the end of the eight lessons of Chapter 1, the teachers were asked to write a final report about what they thought of the CATs.

To answer the first research question about the use of the CATs, teachers were asked whether they used a CAT as suggested in the teacher guide of the CATs. In case they did not, they were asked to indicate which changes they made and why they made these changes. The changes made by the teachers and their reasons for adapting the CATs were categorized based on the responses of the teachers. The initial categories were formulated by scrutinizing Teacher A's answers related to the CATs she used. For example, Teacher A mentioned that she only used three tasks in CAT-1 because she spent quite some time on this CAT and needed to finish other activities. If a teacher's response did not fit into any of the extant categories, a new category was included. For example, Teacher C gave a different reason for reducing the number of tasks in CAT-1; besides saving time, she thought two tasks were similar and one of these could be removed. In the end, this led to three types of adaptations: changing the number of tasks in the CAT (reducing tasks/adding tasks), changing the moment of using the CAT (after class instead of during class/at another moment in class), or changing the procedures of conducting the CAT (deleting steps/adding instruction). The reasons why the teachers made these changes were divided into the following four categories: shortage of time, redundancy of the tasks in the CAT, mismatch with the objectives of the lesson, and difficulty level of the CAT (too easy/too difficult).

The second research question, about what information the teachers got from a CAT and the use of this information, was answered by asking the teachers whether a CAT provided them with new information about their students. If this was the case, they were asked what new information they got from the CAT. Similar to the way in which the teachers' answers to the first research question were processed, we developed the categories based on the teachers' responses. A distinction was made between the content of the new information and its focus. For the content, we had two categories: unexpected findings regarding the correctness of students' answers and unexpected findings regarding their applied strategies. With respect to the focus of the new information, we had three categories: information about the whole class, information about individual students, and information about the difficulty level of the tasks in the CATs. The teachers were also asked whether they used the new information to give additional instruction, and if yes, what they did with this information. The responses were divided into two types: instruction given during or immediately after the CAT, and instruction given in the next lessons. The reasons for not using the new information to provide additional instruction included the following three categories: shortage of time, satisfaction with students' performance, and having no clear clue how to use the information.

For the third research question, about the teachers' perceived usefulness of the CATs, first, we counted for each CAT how many teachers answered that they would use the CAT in the future or not. Then, their reasons for using or not using it were classified, again based on the teachers' responses. With respect to using the CAT in the future, we identified the following four categories of reasons: the CAT can reveal students' learning, can be used as a teaching activity, can enhance students' engagement, and can be carried out in a feasible way. The three reasons for not using the CAT in the future were as follows: mismatch with what is taught or examined, shortage of time, and satisfaction with students' performance. A further resource for answering the third research question was provided by the final report in which the teachers were asked whether they liked the CATs and what they think about their usefulness.

Before an interview was held, the teacher's lesson was observed and video-recorded by the first author. The purpose of these observations was to check the teachers' self-reported information given in the interviews and on the feedback forms. In case there were discrepancies or when particular information was missing, this was discussed with the teachers, and if necessary, the information in the interviews and feedback forms was corrected.

Finally, in case the CATs required the use of worksheets, the written work of the students was collected and analyzed with the focus on their answers and strategies. This provided us with background information when processing the teachers' responses in the interviews and on the feedback forms.

3. Results

3.1 An overview of the teachers' use of CATs

All teachers used at least 11 out of 13 CATs in their practice. They all made changes in the CATs and did not do exactly what was suggested in the teacher guide of the CATs. The reason for this was that they already had lesson plans for each lesson in Chapter 1 before the first meeting took place. These lesson plans were very detailed. For example, they described the number of exercises the teachers should do in each class and the time it would cost to do these exercises. Because the teachers had already a very clear picture of what they were going to do in class, they had to merge the CATs into their lesson plans. They did this very carefully in order to complete their pre-arranged activities and at the same time benefit from trying out the CATs. One of the changes the teachers reported most often was reducing the number of tasks in the CATs. The teachers considered some tasks in the CATs to be redundant and did not like to repeat a similar task. A second change the teachers often made was carrying out the CATs in a time slot before or after class, like in morning reading sessions or self-study lessons, instead of during class. In this way, the CATs would not take precious time from the mathematics class. The teachers were very concerned about the shortage of time in class. Adding extra instruction or help during carrying out the CATs was a third change often made by the teachers. Also, it was found that the teachers from the same school made the same changes in the CATs, which was not such a surprise because in the meetings the teachers discussed with each other how to use the CATs.

By using the CATs, the teachers could clearly see the students who were not able to answer the questions correctly or those who did this with hesitation. Although the teachers noticed that the questions asked in the CATs focus more on revealing students' mathematical understanding rather than checking their calculation skills, it seemed that they nevertheless paid more attention to the accuracy of the answers than to the strategies used by the students. Moreover, the teachers reported getting more specific information about individual students, especially when students were asked to give answers by showing the cards. While conducting the CATs, the teachers often directly provided explanations or helped students to solve the problems. As Teacher B said, “I cannot continue while leaving half of the students to be unclear about how to solve the problems.” However, besides this direct help, no evidence was found that the teachers used the information gained from the CATs to adapt their instruction in the next lessons to meet the students’ needs. For example, no teacher added extra exercises or organized an extra discussion on findings that came to the fore through the CATs. According to the teachers, the main reason was the shortage of the time. The teachers needed to complete the activities they had planned beforehand for the next lessons. So they did not have time to do additional or adapted instructional activities based on the assessment information. Moreover, the lessons were already full because of the CATs to be carried out.

All teachers agreed that the CATs were helpful to know more about their students’ learning. By asking different questions than those in the textbooks, they knew more about whether students had difficulties. Particularly, the teachers recognized the power of using the red/green cards as a tool to quickly gather information about how many students had difficulties and to engage students. In fact, the teachers liked using the CATs that employed this whole-classroom immediate response format. Moreover, the teachers also acknowledged that the CATs gave them insight in what content and skills their students should learn and how to teach them. In line with this, based on this insight provided by the teacher guide of the CATs, all teachers changed their originally planned instructional activities before using the CATs in class. This was done not only because they thought that what would be assessed by the CATs was important to be taught but also to avoid that the students would perform badly on the CATs. There were even three teachers (Teachers A, C, and D) who intentionally taught the tasks in the CATs before offering them as assessment tasks. A further finding was that two teachers (Teachers A and B) used characteristics of the CATs in their own teaching, such as offering their students a series of ordered problems without asking students to calculate the final answers. These two teachers and a third teacher (Teacher E) also designed their own CATs, in which they asked their students to answer by means of the red/green cards. When the teachers decided whether to use particular CATs in the future, their primary concern was whether the CATs fitted to the topics or objectives of their fixed lesson plans. Practical considerations such as the time it costs, the feasibility, and the tasks' difficulty were also important criteria.

3.2 Results for CAT-1: Identifying the watershed

3.2.1 The teachers' use of CAT-1

According to the teachers' responses in the interviews and feedback forms, they made two types of changes when using CAT-1. First of all, to save time in class and to avoid a repetition of tasks which they considered to be similar, all teachers left out one of the four tasks. For example, the four teachers in School II agreed that it would be better to use all four tasks if they had sufficient time in class. However, since they had only 40 min and the planned activities had to be finished, they had to compress the tasks in CAT-1. To them, it seemed there was no essential difference between Task 2 and Task 3 because the dividends were both three-digit numbers. The other type of change reported by three teachers (Teachers A, D, and E) was that they provided extra help when doing the CAT, such as pointing out, by themselves or by good students, the rule for finding the breaking point. For Teacher A, this was confirmed by the video-recording of her lesson. During the process of conducting the first task in CAT-1, she stopped providing other divisors when she noticed quite a few students did not answer correctly when the divisor was 7. Then she asked one of her best students to explain her way of solving the task and reminded her students to think over what the good student just said before starting Task 2. According to Teacher A, such support or help was necessary and it would not have been useful to continue when students did not understand how to deal with the problems.

Besides teachers' self-reported changes in terms of reducing tasks and adding instruction, it was found from video-recordings that Teachers A and C also made other changes. In Task 1, instead of continuing with 4 as divisor after seeing the students' cards when the divisor was 3, Teacher A stopped to check whether a girl understands the question or not. This short break happened right before the moment when students were supposed to change and show their red cards. Another change was that Teacher A reduced the steps of carrying out the CAT by choosing only some numbers as divisor. For example, in the task with

721 as the dividend, this teacher only selected 2, 4, 6, 7, 8, 9, and 10. She stated in the interview that it was not necessary to use the complete sequence of divisors for all the tasks, because “[i]t is a bit a waste of time.” Although Teacher A’s decision did not make the watershed disappear her changes might have reduced the students’ experience of progressively approaching the watershed and anticipating the moment that the card has to be changed. Teacher C also added activities between the tasks, for example, asking students to explain or discuss their solutions. After finishing Task 1 (35 as dividend), she summarized the underlying rule of solving the tasks:

> The key is comparing the divisor and the number in the tens place of the dividend. The digit of the quotient would be two if the former [the divisor] is not bigger than the latter [dividend]; if not, the quotient would be a one-digit number. (Teacher C in video; translated from Chinese)

Later, Teacher C asked her students to explain what this rule implies for solving the other two tasks. Another finding was that Teacher C was articulating the watershed notion by giving visual support. In addition to speaking out the divisors, she wrote 1 to 9 on the blackboard (Figure 5) and emphasized the divisors corresponding to the green card by drawing an accolade.

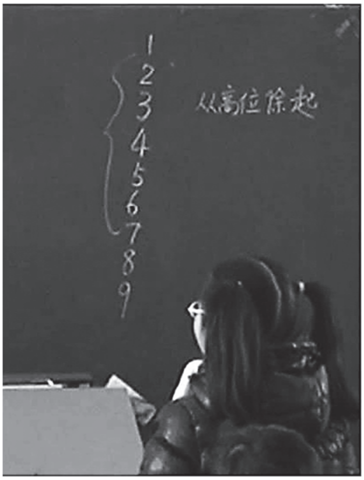


Figure 5. Teacher C’s use of CAT-1 showing the watershed with an accolade for dividend as 721

3.2.2 Information from CAT-1 found and used by the teachers

All teachers agreed that using CAT-1 provided them with new information. However, what information they got was different. All teachers reported that they could see clearly whether their students provided correct answers. Particularly, Teacher F said she only looked at the accuracy of the answers. In contrast, Teachers B and E explicitly emphasized that they also investigated what strategies students used by asking "How did you solve this problem?" Moreover, the information teachers reported finding also differed, regarding their focus: either on the whole class or on individual students. By seeing her students' bad performance in Task 1, Teacher D found that her students had entirely forgotten what they had learned before. Teacher E said that both low and high achievers in her class were interested and participated more than they used to do. CAT-1 also helped four teachers (Teachers A, B, D, and E) to identify particular students having problems. In these cases, the teachers corrected the wrong answers immediately and gave their students some instantaneous help. In the end, all teachers concluded that most of their students could identify the breaking point correctly and that only one or two students hesitated or waited when raising their card.

3.2.3 The teachers' perceived usefulness of CAT-1

All teachers liked using this CAT and five of them would use it in the future. Four teachers (Teachers A, C, D, and E) considered CAT-1 as one of their three most informative CATs. The teachers gave various reasons for finding CAT-1 useful. First of all, it was useful for identifying what difficulties which students have. All teachers noted that CAT-1 was good to elicit students' understanding. One reason that was mentioned for this was that the question asked in CAT-1 was separated from calculating the division. In this way, both the teacher and the students were more focused on understanding. As Teacher A said:

> In general, the exercises given to students ultimately focus on calculation, even if students were asked to make a decision about how many digits the quotient has [Figure 1, exercise 3]. But if [students were] not [asked to] calculate, more attention will be paid to understanding. (Teacher A in interview; translated from Chinese)

The other reason was that unfamiliar questions may better reveal students' understanding. For example, Teacher C mentioned that, in the beginning, CAT-1 seemed difficult to the students since they were not familiar with answering this type of question, and sometimes asking students questions in a different way was helpful to discern whether they understand the essence of a concept or a procedure. In addition, all teachers recognized the advantage of using the red/green cards to quickly find information about individual students. When using the cards, they asked their students to show the cards in a unified way, like holding one card in each hand (green card in the left hand and red card in the right hand) and raising the cards high enough over the head of the student sitting in front. Some teachers found it difficult to remember the students who made mistakes. For future use, they would like to do some registration (e.g., making notes on a seating chart) to have a clearer picture of individual students' performance.

Secondly, four teachers considered CAT-1 to be helpful for their teaching, because this CAT highlighted a necessary building block for being able to carry out a division algorithm. Like what the Teacher D said: "only if students know how many digits the quotient has, are they able to write the 'number' of the quotient in the right column. Therefore, [CAT-1] is supportive for my teaching." Teacher C, who taught two classes but used the CATs in only one of them, explicitly mentioned that the students with experience of CAT-1 made fewer mistakes in exercises than those without. Teacher A liked this technique because it aroused students' interest in learning mathematics and led students to think systematically.

> This technique is very nice. It made students feel that mathematics is mysterious, because things totally change when crossing a number, which raises students' interest to explore and think. Besides, students also benefit from the way in which a kind of orderly thinking is reflected. So if they cannot find the answer, they can start to try from 1. (Teacher A in final report; translated from Chinese)

Besides the two main reasons mentioned above, other reasons for using CAT-1 in the future were increased students' engagement (Teachers B and E) and easiness to conduct (Teachers B and C). Teacher F did not want to use CAT-1 in the future since she thought her students had already learned the knowledge very well.

3.3 Results for CAT-2: Solving division problems without algorithm

3.3.1 The teachers' use of CAT-2

Instead of using CAT-2 during class in lesson 8, as suggested in the teacher guide of the CATs, all teachers conducted it outside the mathematics class, either in a morning reading session or in a self-study lesson. The reason for this change had to do with the format of the CAT. Because the teachers had to check students' responses to CAT-2 after class, during class no immediate help was needed. Therefore, the teachers decided to use all the time during class for activities that required their help and feedback.

According to the teachers' reports, the students were given 10 min at most to solve the four division problems. The checking work by the teachers was partly done immediately after the students handed in their worksheet and partly after the morning reading session or the self-study lesson. All teachers only quickly looked at the student work to get a basic idea about students' performance in terms of correctness, strategies, and mistakes.

3.3.2 Our analysis of the student work of CAT-2

Before discussing the reactions of the teachers, we give an overview of the students' solutions to the first two tasks of CAT-2, based on 189 students' worksheets collected by the teachers. For 468 ÷ 2 (Task 1), 186 students came to a correct answer, and 158 of them provided clear explanations of how they solved it. When zooming in their solutions, it was found that instead of solving this division without using the standard algorithm – as was demanded – more than half of these 158 students basically used the algorithm. Although they noted their solutions in horizontal number expressions, suggesting that they carried out a number of subdivisions based on splitting the dividend, in reality they did a step-by-step processing of digits, similar to the standard algorithm. Therefore, one might wonder whether these students whose work is shown in Table 4 (a) really understood the division operation. A solution that gives a better guarantee for having this insight is using the number values of the dividend by splitting 468 into 400, 60, and 8, making three divisions, and adding the results. Such a solution is shown in Table 4 (b).

Table 4

Types of solutions for 468÷2 and the percentage of students using them

Solution type	Percentage of students[z]	Example
a Digit-based splitting the dividend	60%	4÷2=2 6÷2=3 8÷2=4 46÷2=23 8÷2=4 23+4=234
b Whole-number-based splitting the dividend	40%	400÷2=200 200+30+4=234 60÷2=30 8÷2=4

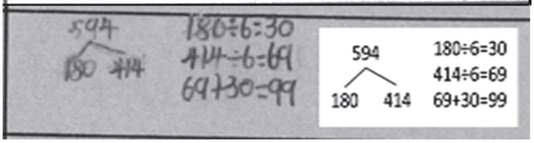
[z] The percentage is based on the number of students who provided a clear explanation of how they solved the problem (N = 158).

However, the real proof of having a good understanding of the division operation is delivered by Task 2, where the students had to solve 594 ÷ 6 without using the division algorithm. The majority of the students, 167 out of the 189 students, could find out the correct result, and 127 students gave their solutions. Approximately three quarters of this latter group stuck to the algorithm either by describing it in Chinese (Table 5 (a)) or notating the algorithm in a horizontal digit-based way (Table 5 (b)).

Yet, while still using a digit-based approach, one tenth of the students were also aware of the number value of the digits (Table 5 (c)), indicating that they had a notion of what is going on in the division. Notwithstanding this, their solution was still based on the standard algorithm. In contrast, some of the students really applied a non-algorithmic alternative for the standard digit-based algorithm: they split the dividend in two or more whole numbers, divided them all, and expressed the subdivisions in horizontal number sentences (Table 5 (d)). Finally, a few students showed their understanding of the division operation by coming up with a strategy in which they made use of 600 divided by 6 (Table 5 (e)).

Table 5

Types of solutions for 594÷6 and the percentage of students using them

Solution type	Percentage of students[z]	Example
a Verbal description of division algorithm	6%	先用百位上的5除以6不够除，用59除以6上9，再用59除以6上9，所以商是99。[y]
b Digit-based splitting the dividend	65%	59÷6=9……5 54÷6=9
c Digit-based splitting the dividend with the answer expressed as a whole number	10%	59÷6=90……5 54÷6=9 90+9=99
d Whole-number-based splitting the dividend	12%	540÷6=90 90+9=99 54÷6=9
		
e Whole-number-based changing the dividend	7%	600÷6=100 600-594=6 100-1=99
		594=600-6 600÷6=100 6÷6=1 100-1=99

[y] Translated from Chinese: Firstly, I used 5 in the hundreds place divided by 6, which was not enough. Then I used 59 divided by 6. I wrote down 9. In addition, I used 59 to be divided by 6, which equals to 9. So the quotient is 99.

[z] The percentage is based on the number of students who provided a clear explanation of how they solved the problem (N = 127).

3.3.3 Information from CAT-2 found and used by the teachers

Initially, all six teachers were unsure about what information they were supposed to find, and three teachers (Teachers C, D, and E) reported that even when they saw the students' responses they were still doubtful. Therefore, when the teachers were asked what new information they gained from this CAT, they summarized what they had observed in the worksheets. Their conclusion was that the majority of the students gave the right answers for most of the division problems. Furthermore, they noticed that most students explained their solutions, that different solutions were brought up by the students, and which tasks were most difficult for them. Thus, the teachers' main concern was whether the students found the correct answers to the division problems but not whether the students could solve the divisions without using the standard algorithm. Nevertheless, the teachers paid some attention to the strategies and they discerned that some students came up with smart ways of doing the divisions.

To be more specific, the teachers concluded that 468 ÷ 2 (Task 1) was not difficult for the students, "because the students could find the right answer." This conclusion indicated that the focus of the teachers was more on the answers than on the strategies. However, the latter was factually what CAT-2 was about. The teachers' focus on answers changed slightly when discussing 594 ÷ 6 (Task 2). Although in this task almost 90 % of the students came up with the correct answer, this time the teachers noticed that most of the students did not find the answer without using the division algorithm. Teachers A and B recognized that the solution of digit-based splitting the dividend with the answer expressed as a whole number (Table 5 (c)) was not what CAT-2 is asking the students. According to these two teachers, such a solution was "seemingly right" but students were "mixing up different strategies and notations." They also made clear that they did not know how to provide feedback to their students. This was also the case when Teachers A and B encountered some students who used the strategy of whole-number-based splitting the dividend (Table 5 (d)), but split the dividend into two or more whole numbers in a rather far-fetched way (for example, 594 is split into 180 and 414).

In the interviews, all teachers also made clear that they did not know how to deal with the information they got from this assessment, although they found it interesting to see their students' thinking. Despite this, they all recognized that a solution in which the dividend was changed, such as using 600 to solve the division 594 ÷ 6 (Table 5 (e)), provided clear evidence of students' understanding of the division operation. Teacher B was surprised that in her class two students, whom she considered as average (or even weak) students, had used such strategy and gave an excellent performance in this task.

3.3.4 The teachers' perceived usefulness of CAT-2

All teachers, except Teacher B, were unsure whether they would use this CAT again, because they did not know how to make use of their students' answers. Nevertheless, they all agreed that it is important for students to solve division problems using different approaches. For example, Teacher A and Teacher C stated that using CAT-2 reminded them that it was not a good idea to put too much stress on practicing algorithms, but they did not know how to train their students to cope with the question in CAT-2. All teachers, except Teacher E, thought it was reasonable that many students did not perform well in CAT-2. After all, students had not previously been trained to solve problems without using the algorithm. In fact, the teachers were not accustomed to ask students such questions. Moreover, they were not used to think about such questions themselves. Teacher F made clear that she never saw such a question and that she also did not think this type of questions would appear in examinations. Teacher B, however, considered this CAT as her second most informative one and was sure to use this CAT in the future.

> [This CAT] expands students' thinking. They are supposed to command how to use the algorithm, but that should not be their only tool. They need to think about the features of particular division problems in order to calculate flexibly, rather than immediately think about the algorithm to solve all problems. (Teacher B in final report; translated from Chinese)

4. Discussion

This small-scale exploratory study was set up to investigate the use of classroom assessment techniques by primary school mathematics teachers in China. Although the six teachers involved in the study did not have earlier experience with these CATs – which is true for the way the content is addressed as well as for the format – they included them rather easily in their lessons by changing them to fit to their pre-arranged lesson plans. Viewed from the perspective of the purpose of formative assessment, it was remarkable that actually no evidence was found that the teachers used the assessment information gained from the CATs for adapting their further instruction, which corroborates the results of the study of Zhao et al. (2006, p. 267), in which they found that "[t]eachers seldom changed their pre-arranged teaching sequence to respond to the needs of their students." In our study, the teachers at most used the assessment information for directly correcting their students' answers thus providing them with instant help in class. In general, the CATs were not used as assessment activities but rather as supplementary exercises. This attitude toward assessment is in agreement with what Cui (2008) and Zhong (2012) found with respect to the classroom practice of Chinese primary school mathematics education: teachers pay more attention to their teaching than to the assessment. This attention paid to teaching is also reflected in the detailed lesson plans Chinese mathematics teachers make (Cai & Wang, 2010; Li et al., 2009), and the fact that the CATs were used as an additional resource for the teachers in refining their pre-arranged lesson plans. This echoes the finding of Cai, Ding and Wang (2014) that Chinese primary mathematics teachers emphasized the design of teaching sequences and questioning based on the study of textbooks and students before teaching. Based on the experiences from our study, we think that an important reason for the teachers not to use the information gained from the CATs to adapt their following lessons is that the teachers gave the highest priority to finishing their already prepared teaching plans. In addition to this, the teachers also reported having difficulty with using the information from the CATs to alter their instruction in the next lessons to meet the current needs of their students.

Despite the fact that the teachers did not use the CATs for informed decision-making about their further teaching, they were quite positive in their evaluation of the CATs as a way to reveal their students' understanding of division. They found the CATs helpful for knowing more about their students' learning and difficulties because the questions differed from those in the textbook. Particularly, the teachers valued the CATs with the red/green cards format for the opportunity they provide to quickly obtain information about students' understanding and engage students. Moreover, the teachers acknowledged that the CATs gave them insight in what content and skills their students should learn and how to teach these.

Although the positive evaluations of the teachers show in a way that CATs can be helpful for Chinese primary mathematics teachers, we also observed that they did not really consider the CATs as a means to assess a deeper level of understanding of division. For example, a teacher's decision in CAT-1 not to present all the possible divisors in a continuous way took away the possibility of the students to discover the breaking point by themselves. By just asking a part of the sequence, the teacher gave away where the watershed is. This resulted in a much less informative assessment because the teacher could not identify whether the students fully understood the relation between the size of the divisors and the number of digits in the quotient. Another example indicating that the teachers may had a different interpretation of the purpose of the CATs is that instead of focusing on examining the students' strategies the teachers were more involved in assessing whether the students found the correct answers. This was clearly the case in CAT-2 where teachers, firstly and mainly, looked at the correctness of the answers and not at whether students could solve the divisions without using the standard algorithm.

Moreover, through CAT-2, also the cultural issue came to the fore. All teachers stated that when they saw this CAT it was not clear for them what information they were supposed to find. The teachers also emphasized that they almost never asked students to answer such questions and almost never thought of such questions themselves. To some extent, this is understandable since East Asian teachers stress the algorithmic side of mathematics and their view of mathematics may result in an emphasis on assessing calculation skills (Leung, 2008). In this respect, there is a difference between teaching division in China and in the Netherlands. Whereas in China much emphasis is put on teaching students the standard algorithm in an early stage, from Grade 2 on, in the Netherlands in Grades 2 and 3, much effort is devoted to give students a good basic understanding of division as equal sharing, making groups, and thinking about the relationship between multiplication and division and stimulate them to use this knowledge to solve division problems. Only from Grade 4 on there is a gradual introduction of the standard algorithm. Therefore, Dutch teachers and students would directly know what to do when they were asked to solve a division problem without using the standard algorithm, but would be lost when asked to determine the number of digits of a quotient before calculating. The positive gain of this clash of educational cultures was that it opened new ways for designing CATs and for assessment problems in general, which was also mentioned by Callingham (2008). It can be very revealing to challenge students with questions that are new for them because they originate from a different educational background and are not prepared by their own textbooks. If the goal of mathematics education is for students to achieve deep understanding of mathematical concepts and procedures, then their knowledge should be able to withstand cultural peculiarities.

Of course, the findings from this explorative study need to be interpreted with prudence, since only a small number of schools (two) and teachers (six) from one district in Nanjing, China, were involved. Whether these teachers' experiences were representative of other teachers in China is something that remains to be investigated.

5. Conclusions

In this study, we explored the use of classroom assessment techniques (CATs) with six Chinese mathematics teachers in primary school. It was found that the teachers could easily include CATs in their daily practice by changing them to fit to their pre-arranged lesson plans. By conducting the CATs, the teachers got new information about their students' learning. In particular, most teachers liked using the CATs with the red/green cards format since they provided quick information of students' understanding. The teachers used this information to give their students, during or after carrying out the CATs, instant help to find the correct answers when the students did not succeed in solving the tasks. However, surprisingly, 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 using the CATs as assessment activities, the teachers often included the CATs in their teaching as extra exercises. Based on the teacher guide of the CATs which gave them insight in what content and skills their students should learn, all teachers adapted their instruction before they conducted the CATs. Some teachers even taught the CATs in advance to avoid their students performing badly on them, which indicates that the teachers did not see in the first place using assessment tasks to figure out what their students can do by themselves. So, formative assessment carried out by teachers to collect information about their students' learning in order to adapt their teaching to their students' needs, which is widely accepted to be a crucial aspect of education, is not so selfevident as one might expect. Our findings indicate that the occurrence of formative assessment in the classroom practice cannot be taken for granted and that the idea of formative assessment by teachers may not always be in line with their prevailing view on teaching. Nevertheless, the teachers valued the CATs as a way to challenge their students with questions that were not completely prepared by textbooks. In addition, they learned from the CATs about what content and skills their students need to learn and how to teach these. Furthermore, they acknowledged the feasibility of using CATs and their potential to engage students. In conclusion, the results suggest on the one hand that CATs can be helpful for Chinese primary mathematics teachers, but on the other hand, our study also provides some evidence that using CATs, as an approach to formative assessment, to make informed and adequate decisions about further teaching, can be a real challenge. This explorative study indicates that more research is necessary into the use of formative assessment in the context of Chinese primary mathematics education.

Acknowledgements

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. All opinions are those of the authors and do not necessarily represent the views of the CSC or NWO. The authors thank Prof. dr. Lianhua Ning in Nanjing Normal University, China, for helping to contact schools, and all the teachers involved in this study for their cooperation and contribution.

References

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability, 26*(2), 153–176.

Berry, R. (2011). Educational assessment in Mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: policy and practice* (pp. 49–61). Netherlands: Springer.

Black, P. (2013). Formative and summative aspects of assessment: theoretical and research foundations in the context of pedagogy. In J. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 167–178). London: SAGE Publications.

Black, P., & Wiliam, D. (1998a). Inside the black box. Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

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

Brown, G. T., 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., Ding, M., & Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265–280.

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*(3), 265–287.

Callingham, R. (2008). Perspectives gained from different assessment tasks on Chinese and Australian school students learning mathematics. *Evaluation & Research in Education, 21*(3), 175–187.

Chen, G. (2006) "Jiti beike" bianxi [An analysis of the practice of "collective preparation of lessons"]. *Journal of the Chinese Society of Education, 9*, 40–41.

Cui, Y. (2008). Jiaoshi ying xian xuehui pingjia zai xuexi shangke [Teachers need to learn how to assess before to learn how to teach]. *Jichu Jiaoyu Kecheng, 11*, 55.

Harlen, W. (2005). Teachers' summative practices and assessment for learning — tensions and synergies. *The Curriculum Journal, 16*(2), 207–223.

Harlen, W. (2007). *Assessment of learning*. London: Sage.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420.

Jiangsu Education Publishing House. (2005). *Sujiaoban jiaokeshu (xiaoxue shuxue sannianji xiace)* [Sujiaoban Textbook (Mathematics textbook for Grade 3 in primary education, Volume 2)]. Nanjing: Author.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: minute-by minute and day by day. *Educational Leadership, 63*(3), 18–24.

Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development, 9*(2), 265–283.

Leung, F. K. S. (2008). In the books there are golden houses: mathematics assessment in East Asia. *The International Journal on Mathematics Education, 40*(6), 983–992.

Leung, F. K. S., Graf, K. D., & Lopez-Real, F. J. (Eds.). (2006). *Mathematics education in different cultural traditions—a comparative study of East Asia and the West*. New York: Springer.

Li, J., & Zhao, W. (2011). "Jiti beike": Neihan, wenti yu biange celue [Collective lesson planning: Meaning, problems and strategies for change]. *Journal of Northwest Normal Univerisity (social sciences), 6*, 73–79.

Li, Y., Chen, X., & Kulm, G. (2009). Mathematics teachers' practices and thinking in lesson plan development: a case of teaching fraction division. *ZDM Mathematics Education, 41*(6), 717–731.

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 biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

OECD. (2012). *Education at a Glance 2012: OECD Indicators*. Retrieved from http://www.oecd.org/ edu/EAG%202012_e-book_EN_200912.pdf.

Popham, W. J. (2000). *Modern educational measurement: practical guidelines for educational leaders*. Needham, MA: Allyn and Bacon.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

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 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. (2017). *Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement*. Manuscript submitted for publication.

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

Wiliam, D. (2011b). *Embedded Formative Assessment*. Bloomington, IN: Solution Tree.

Yu, G., & Jin, Y. (2014). English language assessment in China: policies, practices and impacts. *Assessment in Education: Principles, Policy & Practice, 21*(3), 245–250.

Zhang, D. (2009). *Xiaoxue ketang jiaoxue xingchengxing pingjia wenti yu duice yanjiu* [The problems and solutions of formative assessment of classroom teaching of primary school] (Master's thesis). Available from China National Knowledge Infrastructure.

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. (2017a). 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. (2017b). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education*, Advance online publication.

Zhong, Q. (2012). Ketang pingjia de tiaozhan [The challenges of classroom evaluation]. *Global Education, 1*(10), 10–16.

Authors' contributions

This paper was a collaborative work of the three authors. XZ, MHP, and MV all participated in design of the study. XZ carried out the data collection in Nanjing, China, and was responsible for interpreting the Chinese resources. XZ, MHP, and MV analyzed the data and drafted and revised the manuscript. All authors read and approved the final manuscript. Competing interests The authors declare that they have no competing interests.

Chapter 5

Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques

Xiaoyan Zhao[a], Marja van den Heuvel-Panhuizen[a,b,c], & Michiel Veldhuis[b]

[a]*Freudenthal Institute, Faculty of Science, Utrecht University, the Netherlands*

[b]*Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*

[c]*Nord University, Norway*

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2017). *Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques*. Submitted for review.

Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques

Abstract

This intervention study investigated what insights Chinese primary mathematics teachers gained into their students' mathematical understanding from using classroom assessment techniques (CATs). CATs are short teacher-initiated targeted assessment activities proximate to the textbook, which teachers can use in their daily practice to make informed instructional decisions. Twenty-five third-grade mathematics teachers were involved in a two-week program of implementing eight CATs focusing on multiplication of two-digit numbers. When teachers referred in their reflections to the mathematics 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 insights. About half of the teachers gained insights into their students' mathematics understanding from using at least three of the CATs. However, no significant effect was found of teachers' gained insights on students' change in mathematics achievement.

Keywords: China; Classroom assessment techniques; mathematics education; multiplication; primary school

1. Introduction

1.1 Assessment in the hands of teachers

Any instructional decision making – so in fact any teaching – requires in one way or another information about students' learning (Shavelson, 1973). The more reliable and valid this information is, the better teachers can find a foothold for these decisions. For generating such information many approaches are possible, ranging from standardized externally developed tests to teacher-made assignments. Contrary to the low reliability that, in the past, has been attributed to teachers' judgements of students' performance (e.g., Parkes, 2013), nowadays, assessment in the hands of teachers aimed at gaining insights into their students' progress is highly valued and seen as crucial for adapting the teaching to the students' abilities and needs. Teachers are also considered to be in a good position for collecting information about their students' learning (Harlen, 2007). 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 (Shepard, Penuel, & Pellegrino, 2017). When the assessment focuses on figuring out what students know, or what difficulties students have, for the purpose of making decisions about further instruction, it is considered as formative assessment. Formative assessment in which the teacher has the lead is often referred to as classroom assessment (e.g., Black & Wiliam, 1998; Brookhart, 2004; De Lange, 1999; Shepard, 2000; Stiggins & Chappuis, 2005; Wiliam, 2007).

What information can be collected through classroom assessment depends largely on what assessment activities are conducted. Helpful assessment activities are those which offer teachers a window into the students' thinking to uncover their mathematical conceptions and skills (Wiliam, 2007). Therefore, much attention is being paid to gaining knowledge about how mathematics teachers can improve their assessment activities so as to acquire adequate information about their students' development (e.g., Schoenfeld, 2015). Research has shown that using various oral questioning strategies and written tasks, and then analyzing students' responses, offers mathematics teachers opportunities to reveal their students' understanding (Lin, 2006;

Schoenfeld, 2015). In particular, challenging students with open-ended problems enables teachers to diagnose students' understanding and reveal their ways of problem solving (Lin, 2006; Panizzon & Pegg, 2007). Other measures to make assessments by teachers more informative are using rubrics (Gallego-Arrufat & Dandis, 2014) or concept maps (Jin & Wong, 2015) as frameworks for analyzing students' responses. Both measures were found to assist teachers in identifying gaps in their students' understanding of the particular mathematical topics under investigation.

1.2 Assessment techniques

Assessment techniques are another form of assessment activities by which mathematics teachers can gauge what their students do know and do not know, in order to adjust their teaching to their students' needs. The use of assessment techniques by Leahy, Lyon, Thompson, and Wiliam (2005) can be characterized as short, feasible, and often well-known activities of the teachers, which are fully embedded in their teaching practice. Also, several other researchers and educators (e.g., Andersson & Palm, 2017; Keeley & Tobey, 2011; Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2017; Wiliam, 2011; Wylie & Lyon, 2015) have investigated such assessment techniques.

Wiliam and his colleagues (Leahy et al., 2005; Wiliam, 2011) investigated a large number of assessment techniques to support primary and secondary teachers' formative assessment practice aimed at making instructional decisions for direct use or for decisions regarding their later teaching. Not surprisingly, again, asking questions turned out to be very helpful. However, new was that different types of questioning were distinguished for different moments in the lesson. At the beginning of a teaching sequence range finding questions were used to find out students' previous knowledge (e.g., "how many fractions can you find between 1/6 and 1/7?"; see Leahy et al., 2005). During a lesson hinge point questions were used to indicate the direction of remainder of the lesson (e.g., six polygons are shown and the students are asked to indicate how many lines of symmetry each polygon has; see Wiliam, 2011). Finally, exit pass questions, asked before students were leaving the classroom, were meant to make decisions about the next lesson. Furthermore, to allow all students to answer at the same time, Wiliam and his colleagues suggested the use of ABCD cards and mini whiteboards (Wiliam, 2011). Then, when a question was asked all students could show their answers by holding up a card or writing their answer on the white board.

Similar assessment techniques were also discussed by Keeley and Tobey (2011), who consider these techniques as useful to give insight into students' factual, conceptual, and procedural mathematical understanding for a broad range of mathematics teachers from Kindergarten to Grade 12. In Andersson and Palm's (2017) study it came to the fore that the primary mathematics teachers involved paid most attention to those assessment techniques which helped them best to collect information about their students' knowledge and skills. A similar finding was revealed by Wylie and Lyon (2015) who did research with mathematics and science teachers in high school and found that the most used assessment techniques were asking questions, organizing classroom discussions, and using written tasks.

Characteristic of the aforementioned studies on assessment techniques is that the techniques are all general in nature. They can be applied in any subject and in any mathematical topic. When teachers are provided with such examples of assessment techniques, it can happen, as was found by James and McCormick (2009, p. 976), that some teachers understand the "spirit" behind the assessment techniques and thus are able to adapt them to their teaching, but that others just catch the "letter" of them and carry them out in a ritualized and mechanistic way. This may be the result of providing teachers with assessment techniques that are not directly related to the content the teachers are teaching at that moment. To avoid this, and to have assessment techniques that can generate indications for further instruction, the techniques should be content-dependent.

A study in which this content-dependent approach was chosen is that of Phelan and her colleagues (Phelan, Choi, Niemi, Vendlinski, Baker, & Herman, 2012; Phelan, Kang, Niemi, Vendlinski, & Choi, 2009). The aim of their study was supporting teachers to assess students' learning in pre-algebra. To find out what had to be assessed, an expert panel was organized to map algebra knowledge and its prerequisites. Such a map was used to design the questions and tasks that could provide teachers with the necessary information. This approach turned out to be rather successful and apparently had a positive impact on students' learning (Phelan et al., 2012).

To make the assessment even closer to the teaching at hand, Veldhuis and Van den Heuvel-Panhuizen (2014a, 2017) took the textbook used by the teachers as a starting point. They designed brief and targeted activities, called classroom assessment techniques (CATs), that teachers could make use of in their daily practice to reveal information about students' learning of a particular mathematical concept or skill. In designing the CATs, decisions had to be made about what content needed to be assessed, and what questions should be asked in order to give clear insights into the students' understanding. Designing these questions does not mean just repeating the tasks that are in the textbook. It requires another approach. The assessment should reveal the understanding behind the tasks in the textbook and should include questions that can give teachers access to a deeper level of students' skills and understanding by presenting the content from a different perspective and in an unfamiliar way. In addition, the assessment should have a format that supports teachers to gather the students' information efficiently and effectively and that makes the assessment feasible to carry out. The two main formats that Veldhuis and Van den Heuvel-Panhuizen (2014a, 2017) used for their CATs were red/green cards and worksheets. Having the students responding to a question by holding up a red or a green card the teacher can quickly gain information about the group as a whole. The worksheets, mostly containing a few problems on a specific mathematical concept or skill, are meant to provide teachers with more information on individual students' strategy use. Also in these Dutch studies, positive effects of using the CATs have been found on the students' mathematics achievement (Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2017).

1.3 A new approach to assessment in mathematics education in China

In mathematics education in China, there exists a deeply-rooted examination culture (Zhang & Lee, 1991). External examinations at school level and teacher-made end-of-chapter tests at classroom level used to be the main aspect of teachers' assessment activity (Li, 2000). In 2001, the Ministry of Education (MoE, 2001a) formally launched a new approach to assessment aimed at improving teaching and learning. Since then, mathematics teachers are encouraged to get a comprehensive understanding of students' learning by employing various approaches, for example, written tests, oral tests, open questions, activity reports, observations, interviews, exercises in and after class, and portfolios (MoE, 2001b, 2011). However, Cai and Wang (2010) found that

Chinese mathematics teachers in primary education put more emphasis on providing information to students than on getting information from students. And when taking action to understand students' thinking, teachers are more likely to do so before lessons than during or after lessons (Cai, Ding, & Wang, 2014; Zhu, Yu, & Cai, 2018). Moreover, when teachers plan their teaching, textbooks serve as the main source rather than findings from assessing their students' learning (Cai & Wang, 2010; Li, Chen, & Kulm, 2009). Also, for making decisions about teachers' assessment activities the exercises provided in textbooks have an important role. Yet, such exercises are maybe more suitable for summative assessment than for classroom assessment (Liu, 2012). According to Liu (2012), this may lead to teachers focusing on assessing the result of learning, namely what basic knowledge and skills students have acquired, instead of assessing how students developed their mathematical thinking during the learning process. Furthermore, studies have revealed that only very limited attention has been paid to improving teachers' assessment practice to get more information about students' learning (Gu & Gu, 2016; Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2017, *Chapter 3 of this thesis*).

Taking into account the promising international findings about the use of assessment techniques, we explored whether this approach to assessment could assist Chinese primary mathematics teachers in their assessment practice. As a sequel to studies carried out in the Netherlands (Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2017), we investigated the use of classroom assessment techniques (CATs) in China. In a pilot study (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2016, *Chapter 4 of this thesis*), six third-grade mathematics teachers of two primary schools in Nanjing, China participated. The focus in this pilot study was on assessing the topic of division, in particular three-digit numbers divided by a one-digit number. In line with the Dutch study (Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2017), CATs in the Chinese pilot study were also based on a textbook analysis, and formulated in such a way that they were not just a repetition of what is in the textbook. In this way, the CATs might give teachers access to a deeper level of students' skills and understanding. It was found that teachers recognized that it can be very revealing to challenge students with questions that are not completely prepared by the textbook. Also, they appreciated the use of red/green cards for providing quick information. In general, teachers were positive about the CATs as a way to reveal their students' learning in an effective and efficient way.

1.4 The present study

Based on these experiences, we set up an intervention study to investigate whether this positive finding holds for a larger group of teachers, and for a different mathematical topic. In particular, we wanted to delve deeper into the insights Chinese primary mathematics teachers can gain from carrying out the CATs. This means that our focus was on investigating whether the teachers, through using the CATs, could acquire knowledge about their students' learning that they did not have before. Therefore, the first research question of the present study was: *What new insights can Chinese primary mathematics teachers get about their students' understanding of mathematics by using CATs?*

Because through the pilot study we already had experiences with the Chinese mathematics curriculum in Grade 3, we chose to do the present study also in this grade. However, to extend our knowledge about the use of CATs in Chinese mathematics classrooms, we changed the topic of investigation. We stayed in the domain of number and operations, but instead of on division the focus in the present study was on multiplication, in particular on what insights teachers can gain from using CATs to assess students' understanding of multiplication of two-digit numbers.

Furthermore, as using CATs in Chinese primary mathematics education is a quite new topic, we also explored the relationship between teachers' use of CATs and their students' achievement. Therefore, our second research question was: *Are insights into student learning the teachers gained from using CATs related to changes in their students' mathematics achievement?* With respect to this question, our hypothesis was that teachers' use of CATs can offer them opportunities to become more informed about their students' understanding and skills in mathematics learning, and that this information may allow teachers to better adapt their teaching to their students' needs, which in turn is expected to lead to improved student achievement.

2. Method

In this short intervention study, Chinese teachers were asked to integrate the use of a number of CATs into their regular teaching of multiplication during the first two weeks of the second semester of Grade 3. Teachers were informed about the CATs through a teacher guide and two researcher-led meetings. Data on how teachers used the CATs and what insights they got from them were gathered through feedback forms and a teacher-written final report. Pre-/posttest data of the students' mathematics achievement were collected by means of district mathematics tests.

2.1 Participants

To answer the research questions, we decided for practical reasons to set up the study in Nanjing. The recruitment of schools and teachers was started from October, 2014. We contacted three local teaching research offices, which are responsible for inspecting the educational quality of the schools and for providing professional development to primary school teachers in their administrative districts. One of these offices volunteered to participate. In order to include various schools in terms of the school's reputation, educational quality, and location, nine out of forty primary schools were selected by this local teaching research office. Of these nine schools, the Grade 3 mathematics teachers and their students took part in the study. Our sample thus consisted of 25 teachers and their students in 25 classes. In all classes the same textbook series was used, namely the Sujiaoban (苏教版) textbook (Jiangsu Phoenix Education Publishing House, 2014).

2.2 Multiplication of two-digit numbers

For developing the CATs, we first investigated when and how the topic of multiplication of two-digit numbers was addressed in the Sujiaoban textbook. We found that this topic was dealt with in the first chapter of the book meant for the second semester of Grade 3. This chapter covers nine lessons taught in around two weeks, consisting of so-called new lessons and revision lessons. A new lesson mostly starts with a new type of problem presented as a context problem, followed by the corresponding bare number problem. Then, examples are given of how to solve this problem type, and finally, exercises are offered to practice this. A revision lesson generally includes exercises for rehearsing and discussing what the students have learned in earlier lessons. The main content components addressed in this chapter, include, among others, multiplication with multiples of ten and the structure of the multiplication algorithm.

Multiplication with multiples of ten is presented in a new lesson and starts with a context problem in which Uncle Li is sending 10 boxes of bell peppers, with 12 peppers in each box. This context problem is followed by the corresponding bare number problem. The students need to find out how many peppers are sent in total. The textbook shows that one method of solving 12 × 10 is to make use of 12 × 1. By seeing both multiplications with their answers the students become acquainted with the strategy of using an analogous problem, that is, using a problem of which the answer is known or easy to calculate to find the result of an analogous problem. Hereafter the textbook provides three sets of exercises with multiplications of multiples of ten: 16 × 1 =, 16 × 10 =; 70 × 6 =, 70 × 60 =; 5 × 40 =, 50 × 40 =.

In the next new lesson, the structure of the multiplication algorithm is introduced. Here, special attention is paid to how the result of the multiple-of-ten part of the multiplication is notated, namely without writing down a zero and leaving the one-position empty (see Figure 1). This structural understanding is further supported by exercises in which the students are provided with an empty structure of the multiplication algorithm that they have to fill in (see Figure 2). In addition, the students have to explain what they need to calculate in each step.

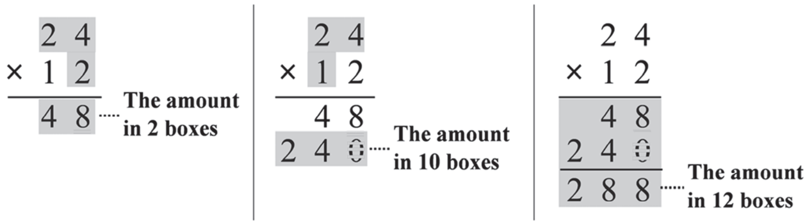


Figure 1. Structure of the multiplication algorithm.

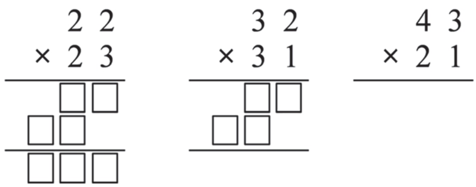


Figure 2. Exercises with empty multiplication structure.

In the subsequent new lesson, the students are prompted to further strengthen their understanding of the structure of the multiplication algorithm. To achieve this, the textbook offers only the start of the algorithm for 24 × 53 (see Figure 3). The students have to complete the remaining steps of the algorithm. Right after this, the textbook provides a description in words of the steps to be taken when carrying out the algorithm of multiplication of two-digit numbers. The students are told to first choose the digit in the ones place of the lower number to multiply the upper number, and then do the same for the digit in the tens place of the lower number. Hereafter, for every calculated product, they have to write the last digit of the product in the same column as the digit chosen from the lower number. Finally, the students need to add the two products.

$$\begin{array}{r} 2\;4 \\ \times\; 5\;3 \\ \hline 7\;2 \end{array}$$

Figure 3. Multiplication algorithm of which the start is given.

2.3 CATs for assessing multiplication of two-digit numbers

To provide teachers with a tool for getting insights in their students' learning and actionable clues for their next steps in teaching, we developed eight CATs: five CATs in the format of the red/green cards, three in a worksheet format. Each CAT contains two similar tasks for doing two assessments, if necessary. Appendix C contains a complete overview of the CATs. Here, exemplarily, we explain three CATs in detail. Two are meant for assessing multiplication with multiples of ten (CAT-1) and the structure of the multiplication algorithm (CAT-3). Finally, near the end of the chapter, when students have learned the multiplication algorithm of two-digit numbers, whether their understanding goes beyond mechanically carrying out the algorithm is assessed (CAT-8).

2.3.1 CAT-1: Family problems

Multiplications with multiples of ten are often considered as rather easy problems. Solving 12 × 10 by thinking of the analogous problem 12 × 1, and adding a zero is not hard. However, understanding why this simple adding of a zero works, is something else. To really grasp the content component of multiplication with multiples of ten, a deeper understanding of the ten-based number system is necessary. Just being able to put a zero at the end of a result, in the case of problems within the number range of two-digit numbers (e.g., using 70 × 6 to find 70 × 60), does not guarantee that the students comprehend this content component of multiplication. Therefore, using the exercises in the textbook in which the numbers are below one hundred, has limited value for assessing whether students truly understand multiplication with multiples of ten. Students have learned to add one zero in the chapter, and in the assessment based on these problems, they have to add one zero too. Students can pass this test by carrying out mechanically what they have practiced. To learn more about students' understanding, we developed a CAT in which the scope went beyond the two-digit number range. If students understand the ten-based number system, they can use the analogy strategy also for a broader collection of problems.

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 4. CAT-1: Family problems.

CAT-1 (Figure 4) has the green/red card format and starts with the multiplication 97 × 8 of which the answer is given. Then several other, related, multiplication problems follow, such as 970 × 8000. These problems are at first sight not easy to solve by mental calculation. In the CAT, students are not asked to solve these problems but only whether they think they are able to solve them. They show their answer by raising the green ("Yes") or the red card ("No"). By inspecting the green and red cards the teacher gets an immediate overview of students' responses. In this way, he/she observes whether students' understanding of multiplication with multiples of ten goes beyond mechanically adding one zero and whether they see the analogy and think they can make use of it.

2.3.2 CAT-3: Breaking down a multiplication

Knowing how an algorithm is built up can help to use it. Therefore, in the chapter much attention is paid to the structure of the multiplication algorithm.

Students were explained how the results of multiplications with multiples of ten are notated, they had to fill in an empty structure of the multiplication algorithm, and they were taught how to carry out the algorithm step-by-step. However, being able to write down the algorithm perfectly and even arriving at the correct answer does not necessarily mean that students understand what they are doing and understand the structure of multiplications with two-digit numbers.

24×53 means that you have to calculate

a)	3×4	and	3×20	and	__×4	and	__×20
b)	4×__	and	4×__	and	__×3	and	20×50
c)	20×50	and	20×__	and	4×50	and	__×3

Figure 5. CAT-3: Breaking down a multiplication.

CAT-3 (Figure 5) has a worksheet format and is meant to give teachers an extra opportunity to assess whether their students can identify the components of a multiplication and understand what is behind the algorithm. In this CAT, the same numbers are used as in the textbook, namely 24×53. However, now students have to unravel this multiplication instead of carrying it out. By using distributive and associative properties this can lead to four sub-multiplications, namely 3×4, 3×20, 50×4, and 50×20, or in any other order. The teacher hands out the worksheet and checks students' responses after class and uses the gained information for decisions about further instruction.

2.3.3 CAT-8: Solving problems without algorithm

The exercises provided in the chapter are mainly on solving multiplication problems, with or without context, by using the algorithm. By the end of the chapter, it is expected that most students are quite able to correctly perform the algorithm. However, after lots of practice, it could happen that students carry out every step of algorithm perfectly but merely in a mechanical way. CAT-8 (Figure 6) has a worksheet format and assesses whether students really understand what a multiplication means and thus also what the algorithmic procedure actually implies.

Here you see a multiplication problem. You have to solve it WITHOUT using the algorithm. Write down how you solved it.

	In this way I solved the problem	Answer
59×62		______

Figure 6. CAT-8: Solving problems without algorithm.

In this CAT students are required to solve multiplication problems without using the algorithm. The main idea is that when students cannot solve a multiplication problem without using the algorithm they will probably not have a sufficient understanding of multiplication, which might put them in trouble when learning to solve more complicated multiplication problems with, for example, three-digit numbers or decimal numbers. Examining the worksheets after class offers the teacher clues about whether and what instructional supports students need before finishing this chapter.

2.4 Data collection

To familiarize the teachers with the CATs and assist them in using them in class, two two-hour meetings were organized. The first meeting took place before the teaching of the chapter on multiplication started. Each teacher received a package including a teacher guide with the material (PowerPoint slides, red/green cards, worksheets) needed for carrying out the CATs in their teaching of this chapter. During the meeting, some general information about classroom assessment was presented and the CATs that could be used in the coming week were discussed. The second meeting began with sharing experiences of the teachers in using the CATs. Hereafter, the CATs for the second week were discussed.

To know what insights into students' learning teachers might gain, they were asked to fill in feedback forms and write a final report. Teachers filled in the feedback forms every time they used a CAT and wrote the final report after finishing the chapter. Specifically, on the feedback form, the teachers filled in whether using the CAT helped them gain new information about their students' learning, whether they adapted their further instruction, and what the new information and the instructional adaption looked like. In the final report, among others, the teachers were asked to suggest two CATs to be included in the textbook as assessment exercises.

To look into the relation between teachers' gaining insight and changes in students' mathematics achievement, we used data from district. Due to district regulations, it was not possible to use other tests. The three tests we used were: (1) the end-term test of the first semester of Grade 3 (Pretest), (2) the multiplication chapter test (Posttest 1), and (3) the mid-term test of the second semester of Grade 3 (Posttest 2). These tests were designed and arranged by the local teaching research office, and the teachers administered the tests to their students. The tests had items on the domains of whole number operations, fractions, measurement, and geometry, but they varied in the domains included. Nevertheless, they all contained multiplication problems. Although the number of items per test differed, for each test the maximum number of points was 100. After the students took the tests, the local teaching research office provided the correct answers to the test items and the criterion for grading; the teachers were responsible for grading their students' test papers and calculating the total scores.

2.5 Data analysis

2.5.1 Qualitative data

In order to get an overall picture of the CATs teachers used, we first scanned all the filled-in feedback forms. In the final analysis, the responses of 25 teachers on the feedback forms and final reports were included. To answer the first research question, regarding teachers' insights into students' mathematics learning, we had to identify which teachers got insights when using the CATs. For making these decisions, all teachers' responses were gathered and translated into English. First, all the responses were scrutinized by the three authors separately. Each author identified the information the teachers gained about their students. Together with identifying whether there were indications for evidence of gaining insights, the authors also had to specify why they thought so. The latter was meant to come to criteria for having gained insights. Hereafter, for each CAT, the decisions on whether the teachers' responses showed indications of insights, and the reasons for making these decisions, were compared and discussed among the three authors. In some cases, whether there was evidence of teachers gaining insights was obvious. For these teachers' responses, it was easy to arrive at the same decision. When the decisions differed, they were discussed at length until 100% agreement was reached. Then, we checked all of our decisions again which led to some changes of our earlier decisions. During this checking process, we also finalized the formulation of the criteria for indicating gaining insights.

In the end, this resulted in four unique criteria:

1. *Referring to the mathematical content the CAT is supposed to assess* (see Sections 2.3.1-2.3.4). For this, teachers can use their own words, or give a clear description of the purpose of the CAT by using (partly) the wording that appeared in the teacher guide. However, this criterion is not met when teachers only refer in general terms to the CAT and do not mention the mathematics that is assessed.

2. *Providing specific information about students*. This includes mentioning the proportion of students showing a particular performance on the assessed content or describing the difficulties students encountered with this content.

3. *Describing the novelty of the gained information about students*. This means that teachers learn something "new", "out of expectation", "surprising", or "that was not known before" about students' understanding of the assessed content.

4. *Explaining an instructional adaptation fitting to the findings from the CAT*. Such an instructional adaptation has to correspond to the information about the assessed content as revealed by using the CAT; general phrases like "providing additional exercises" or "give extra instruction" are not sufficient.

Table 1

Examples of the teachers' responses about CAT-1 Family problems, and whether they were qualified as showing having gained insight

Example of teacher's response	Criterion met	Having gained insights
Example 1: Teacher S14 "In my expectation, the vast majority of students can find the correct answer without being disturbed by the increasing number of '0's. However, the reality is that students felt difficulty when the numbers became bigger and bigger [II, III]. Therefore, it illustrated that students were not flexible enough when solving by mental calculation the problem of multiples of ten [I]. Also it means that students lack the ability of reasoning and generalization [I]."	I, II, III	Yes
Example 2: Teacher S09 "The students could not recognize such analogous problems very well [I]. Most students showed their card based on their 'gut feeling' [II]. Only a small proportion of the students could find the rule [II]. As the number of '0's increased, the accuracy went down [II]."	I, II	Yes
Example 3: Teacher S04 "Most of the students could give the answer quickly, but some individual students need help [II]."	II	No
Example 4: Teacher S10 "This CAT helps students to extend the boundary of knowledge. The difficulty level is appropriate. The students showed strong interests when doing the CAT. In general, the CAT is helpful to extend students' knowledge and improve their initiatives."	[no criterion applicable]	No

Showing that one has learned something from doing an assessment is a multifaceted phenomenon. It can be expressed in different ways. Teachers can say something about the performance of their students, can emphasize that they discovered new information in the students' performance, or can discuss their decisions about further teaching. All these responses can give some indication that teachers have learned something from assessing their students with the CATs. Yet, to fully classify a teacher's response as having gained insights from a CAT, a first requirement is that the teacher refers to the mathematical content the CAT is supposed to assess. Just talking about students' performance in general terms is not sufficient. So, our final decision rule to qualify a teacher's response as having gained insights is that it should meet Criterion I and at least one from Criterion II, III, and IV. Based on this decision rule, a final round of checking was carried out by the first author. This resulted in qualifying 57 teachers' responses out of the total 200 possible responses (25 teachers × 8 CATs) as having gained insights. Table 1 provides examples of the qualifications of teachers' responses about CAT-1.

2.5.2 Quantitative data

To answer the second research question, on the relation between teachers' use of the CATs and changes in their students' mathematical achievement, we examined if gaining insights from the CATs is related to students' changes in test scores over time. For this we used the scores on the three district tests. In total, we got results from 981 students. To ensure students' answers to the tests were correctly scored by the teachers, we checked their scoring. Regrettably a number of grading mistakes were found. This finding required that we had to do the entire scoring procedure again. However, checking all the items in the three tests for all the students was not feasible. Therefore, in the end, we decided to randomly select one fifth of the students in each class to check their scores and, if necessary, correct them. This resulted in having double-checked scores of 198 students for the final analysis on the relationship between teachers' having gained insights and changes in students' mathematics achievement.

Descriptive analyses were first conducted for the three tests (*M, SD*). Since the tests had slightly different domain coverages and raw scores were therefore not directly comparable, standardized scores (z-scores) were calculated. With these standardized scores, we looked at differences in posttest scores for different groups of students that were based on the number of CATs for which their teachers showed clear evidence of having gained insights. More specifically, we used analyses of covariance in which we entered standardized posttest scores as dependent variable, standardized pretest scores as covariate, and insight group as fixed factor.

3. Results

In the end, 193 responses (white cells in Table 2) about teachers' use of the CATs were collected, three teachers (H02, H04, N05) did not use all the CATs (black cells in Table 2), so for some CATs there were no responses.

3.1 Teachers gaining insights from using the CATs

In 57 responses (30% of all responses), clear evidence of teachers gaining insights was identified (the white cells with "✓" in Table 2). Every CAT was carried out by at least 23 teachers. Every time at least 5 teachers gained insights into their students' understanding of multiplication of two-digit numbers. CAT-8 appeared to be the most informative CAT for teachers; for this CAT, there were eleven teachers who showed evidence of gaining insights.

According to what the teachers reported, twenty-two teachers used the eight CATs, one teacher (H02) used seven CATs, and two teachers (H04, N05) used only five CATs. To what extent a teacher gained insights into students' learning differed between the teachers. Five teachers reported to gain insights from more than half of their used CATs: they formed the *High Insight* group. For example, Teacher H01 gained insights into her students' learning from seven CATs and Teacher H04 gained insights from every of the five CATs he used. In contrast, in six teachers' responses, no clear evidence of insights was identified no matter how many CATs they used. These teachers made up the *No Insight* group. The remaining fourteen teachers reported carrying out eight CATs but seemed to gain insights from not more than three CATs; they constituted the largest group, the *Some Insight* group. To give more information about the specific insights the teachers reported to have gained into their students' mathematical understanding by using the CATs, their responses about three CATs (1, 3, and 8) are described in the following.

Table 2

Overview about whether clear evidence of teachers (n = 25) gaining insights into their students' understanding of multiplication of two-digit numbers by using the CATs was identified [a]

Insight	Teacher	CAT								
		1	4	7	2	5	6	3	8	Total of ✓
High Insight	H01	✓	✓	✓		✓	✓	✓	✓	7
	H02		✓	✓	✓	■	✓	✓	✓	6
	H03	✓			✓	✓	✓	✓		5
	H04	■	■	✓	■	✓	✓	✓	✓	5
	H05			✓	✓	✓	✓		✓	5
Some Insight	S01			✓	✓				✓	3
	S02				✓		✓		✓	3
	S03		✓			✓		✓		3
	S04		✓					✓	✓	3
	S05					✓	✓		✓	3
	S06	✓				✓			✓	3
	S07		✓						✓	2
	S08					✓	✓			2
	S09	✓						✓		2
	S10								✓	1
	S11							✓		1
	S12							✓		1
	S13				✓					1
	S14	✓								1
No Insight	N01									0
	N02									0
	N03									0
	N04									0
	N05		■			■		■		0
	N06									0
Total of ✓		5	5	5	6	8	8	9	11	57

[a] Black cell means the teacher did not use the CAT; empty white cell means the teacher used the CAT but no clear evidence was identified that the teacher got insights by using the CAT; white cell with "✓" means the teachers used the CAT and clear evidence that the teacher got insights by using the CAT was identified.

3.1.1 Insights from using CAT-1: Family problems

All teachers, except one (H04), returned a response for CAT-1. In five teachers' responses, clear evidence was found that they gained insights about whether, and to what extent, their students understood multiplication with multiples of ten beyond the two-digit number range. Four of the teachers who gained insights referred to the mathematical content CAT-1 is supposed to assess. They reported either about students' analogous thinking (H03, S09) or about students' flexible use of known rules of multiplication with multiples of ten (H01, S14). Teacher S06 described that CAT-1 aims to assess "a method" that "makes use of the given problems" and that shows its "advantage when the number of '0's increases." In their responses, the five teachers dealt with their students' performance when the numbers in the problems contained more than two digits. For three teachers (H01, S06, S14) this came down to reporting that most students were able to deal with this and that just a few students "could only solve [until] 97×80 and 97×800" (H01). Three teachers (H03, S09, S14) reported that fewer students provided correct answers when the numbers became bigger and that "only a minority of the students could determine the rule" (S09). This came as a surprise to one teacher (S14), who explained that "in my expectation, the vast majority of students would find the correct answers without being disturbed by the increase in the number of '0's." Only two teachers (H03, S06) mentioned how they would adapt their instruction; they were going to include analogous problems for students to practice with.

In nineteen teachers' responses, the mathematical content assessed by CAT-1 was not mentioned. When teachers described students' performance, it was in very general terms. Some teachers reported to be satisfied with their students' performance, like students were able to calculate "according to the given characteristics" (S05), other teachers pointed out their students' shortcomings in understanding "the rule" (H02, H05, S01). Similarly, when teachers wrote about their instructional adaptations, they often used general terms, like "more exercises" (S07) or "extra instruction" (N05). Interestingly, two teachers (S11, N04) decided not to adjust their further teaching since they considered the content of CAT-1 to be too similar to what is in the textbook. In contrast, another teacher (N01) provided as a reason for not adapting her instruction that "there is no such type of exercise in the textbook."

3.1.2 Insights from using CAT-3: Breaking down a multiplication

Twenty-four teachers provided a response for CAT-3. Nine teachers were found to have gained insights into whether their students could identify the components of a multiplication of two-digit numbers and understand what is behind the algorithm. Only one of these teachers (S11) reflected on what is assessed in CAT-3 in her own words: “breaking down the multiplication problem into four components is factually the same as showing how the algorithm works to calculate the multiplication of two-digit numbers.” The remaining teachers who gained insights referred to the mathematical content assessed by CAT-3 in terms of “the meaning of multiplication of two-digit numbers” (H03) or “the meaning behind the algorithm” (H01). When describing students’ performance, most teachers (H01, H02, S03, S04, S09, S11, S12) made a clear distinction between students’ understanding of the structure of the multiplication algorithm and students’ ability to apply the procedures. For example, one teacher (S12) reported that most of her students “master the procedure of calculating, but their understanding about how it works is not good enough.” Another teacher (S11) found that “25 out of the 39 students could answer all the blanks correctly.” In contrast, the remaining teachers reported that at least half students were unable to break down 24×53, although they were able to find its result. Two teachers (H01, H04) said this was out of their expectation. Another teacher (S03) also expressed her surprise, as she “thought the students would not even understand the question”, but “the situation in fact was a bit better.” For further teaching, this teacher wanted to pay more attention to help the students to “understand the meaning of each step of performing the algorithm”.

The remaining fifteen teachers did not refer to the mathematical content assessed in CAT-3. Instead, they mainly focused on describing students’ performance. Most teachers reported that their students had difficulties in breaking down the problem, for example “students could only break down 24×53 into $24 \times 50 + 24 \times 3$” (S06), or difficulties in understanding the question, for example “students had never been trained to break down a multiplication problem into four parts” (S05). Overall, students in these fifteen classes did not show good performance on CAT-3. However, about half of the teachers would not adjust their further teaching. The main reason they gave was that CAT-3 was too different from what they taught in class about multiplication of two-digit numbers and it therefore could “disturb students’ thinking” (N01) or “make students confused” (N02).

3.1.3 Insights from using CAT-8: Solving problems without algorithm

All teachers returned a response for CAT-8. In eleven teachers' responses, clear evidence was identified of gained insights into students' capability of solving multiplication problems of two-digit numbers without using the algorithm. In their responses, these teachers referred to the mathematical content assessed by CAT-8, for example about their students' understanding of multiplication, students' flexible use of different solutions, or students having mindset about solving multiplication problems. In particular, these teachers described their students' performance of providing different solutions to solve 59×62. For example, one teacher (H02) found that some students understood the connection between *Lattice multiplication* and breaking down the multiplication. Another teacher (S01) wrote that "a small part of her students could solve the problem by using the distributive property." Contrastingly, another teacher (S05) found that "part of the students thought 'without using the algorithm' meant 'no accurate answer being required', therefore they only made an estimation of the product." Furthermore, two teachers (H01, S10) reported that their students were used to writing down the algorithm when given a multiplication problem, and they did not know how to start now. Only one teacher (H05) really described the novelty of her gained insights. She had expected her students to not be able to solve 59 × 62 without using the algorithm, but in fact many of them used the method of *Lattice multiplication*. Regarding to instructional adaptation, two teachers (H02, H04) valued the "openness and flexibility" (H02) of CAT-8 to "give students more space to think and imagine freely" (H04), and would use it in future teaching.

Of the fourteen teachers with no evidence for insights, most only shortly reported that their students could not solve problems without using the algorithm. Some teachers did refer to CAT-8 as aiming to "develop students' divergent thinking" (S09), to "extend students' learning" (S12), or to "remind students to solve problems in different ways" (S11). Three teachers (S03, S13, N06) seemed not to understand what CAT-8 aims to assess. They thought the assessed mathematical content was students' ability to apply the properties of multiplication, for example "it is difficult for students to understand the distributive property" (S13). Another teacher (S03) made clear that "students are going to systematically learn the properties [in Grade 4]." Seven teachers used general phrases to describe their instructional adaptation, for example “providing additional instruction for those who are able to learn more” (N03) or “providing extra exercises to revise this content” (N01). The other seven teachers would not make any instructional adaptation. Two of them (S11, N04) did this because their students did not have problems in solving the problems, and the remaining five teachers felt using the algorithm was more suitable to solve these problems.

3.2 Relation between teachers gaining insights and changes in student achievement

When looking at students’ mathematics achievement scores, at least two-third of the students got more than 85 (out of 100) points for Pretest (75.8%), Posttest 1 (69.7%) and Posttest 2 (68.2%). Students’ mathematics achievement in the *High Insight* group slightly increased from Pretest ($M_{high} = 87.1$, $SD_{high} = 9.1$) to Posttest 1 ($M_{high} = 87.8$, $SD_{high} = 10.0$), and then decreased to Posttest 2 ($M_{high} = 84.1$, $SD_{high} = 14.2$; Figure 7). Test scores of the students in the *Some Insight* group kept decreasing from Pretest ($M_{some} = 88.7$, $SD_{some} = 8.7$), to Posttest 1 ($M_{some} = 88.2$, $SD_{some} = 8.5$) and then to Posttest 2 ($M_{some} = 85.2$, $SD_{some} = 12.2$). The students in the *No Insight* group had decreasing scores from Pretest ($M_{no} = 90.1$, $SD_{no} = 7.4$) to Posttest 1 ($M_{no} = 88.0$, $SD_{no} = 10.6$), and slightly increasing scores to Posttest 2 ($M_{no} = 88.8$, $SD_{no} = 12.3$).

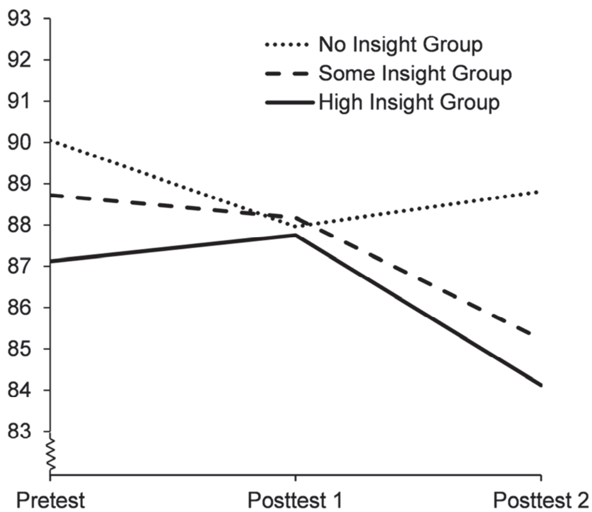


Figure 7. Students’ test scores per insight group from Pretest to Posttest 1, and to Posttest 2.

With the standardized scores, we performed an analysis of covariance (ANCOVA) by putting the Pretest score as covariate and insight group as fixed factor. It was found that the slight differences on the Posttest 1 scores were not statistically significant ($F(2, 194) = .849$, $p = .43$, $\eta_p^2 = 0.009$), the same was found for the Posttest 2 scores ($F(2, 194) = .950$, $p = .39$, $\eta_p^2 = 0.010$).

4. Conclusions and discussion

In this study, we investigated the insights into students' understanding of mathematics Chinese primary teachers gained by using CATs. CATs are meant to give teachers access to their students' deep understanding of a particular mathematical concept or skill, by posing questions that purposely present the mathematical content from a different perspective and in an unfamiliar way. In this way, CATs not only provide teachers with new lenses to observe and understand their students' learning, but also offer them clues to answer the question of what to do next. Although using CATs to assess students' learning, in particular their understanding of multiplication, was quite new for the participating Chinese mathematics teachers, about half (11 out of 25) of them gained insights into students' mathematics understanding from using at least three of the eight CATs. Gaining insights here means that, in a teacher's response to a particular CAT, clear indications are identified that the teacher understands what mathematical content a particular CAT aims to assess, and reflects on what occurs in his/her own classroom by paying specific attention to the students' reactions in critical moments of solving the problems in the CATs or describing the instructional adaptations corresponding to the students' performance on the assessed content as revealed by using the CAT.

Depending on the number of CATs from which a teacher gained insights, they were divided into three groups: the *High Insight*, the *Some Insight* and the *No Insight* group. In line with what was found by James and McCormick (2009), the *High Insight* teachers seemed to be able to understand the idea of the CATs, but the *No Insight* teachers might have only followed the prescribed procedure of CATs and carried them out in class without getting better understanding of students' mathematics learning. Particularly, the teachers belonging to the *High Insight* group favored the revelatory capacity assessment tasks that differed in a specific way from the teaching tasks they would generally provide, whereas the teachers in the *No Insight* group failed to realize such advantages. Several teachers considered CATs to be (too) similar to the textbook and did not want to repeat what they had already taught. In contrast, others emphasized that CATs were too different, or too difficult, compared to their regular teaching. The latter group of the teachers appeared to hold a rather narrow view of what should, and could, be assessed. Despite the differences between teachers in different insight groups, in their responses they all paid more attention to descriptions of students' performance than to possible instructional adaptations. This echoes the finding of Heritage, Kim, Vendlinski, and Herman (2009) that mathematics teachers found it more difficult to decide about next instructional steps from assessment information than to notice students' understanding.

When interpreting these findings, it should be borne in mind that decisions about whether the teachers got insights from using CATs were based on their self-reported data. Further data collection, like observing teachers using CATs in class and directly asking about their gained insights, could shed new light on possibly gathered new information about their students' mathematics learning. Moreover, it needs to be acknowledged that the CATs used in the context of Chinese primary mathematics education so far, including the pilot study (Zhao, et al., 2016, *Chapter 4 of this thesis*), were designed based on one particular textbook series and only involved teachers from the city of Nanjing. Whether Chinese primary teachers who use different mathematics textbook series or who are from different regions can get new insights from implementing CATs remains unclear. Notwithstanding these limitations, the study suggests that Chinese primary mathematics teacher can gain insights into their students' mathematics understanding from using classroom assessment techniques. It would be interesting to explore whether teachers with different assessment profiles (Veldhuis & Van den Heuvel-Panhuizen, 2014b; Zhao et al., 2017, *Chapter 3 of this thesis*) benefit differently from carrying out CATs. For example, recently it has been shown that significant differences exist between expert Chinese primary mathematics teachers and their non-expert colleagues in their perception and reported behavior of understanding their students' mathematics thinking (Zhu et al., 2018).

A second issue we explored is whether teachers having gained insights into their student learning from using CATs was related to changes in their students' mathematics achievement. In our study, no significant effect of having teachers' gained insights on student achievement was found, which is different from some other international studies (Phelan et al., 2012, Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2017). A possible explanation could be the short period of the intervention. Within two weeks, teachers were informed about the purpose of CATs and how to use them, had to incorporate them into their teaching practice, reconcile the new insights about their students with their original understanding, and take actions to adjust their teaching. Another explanation could be the strong ceiling effect on the tests, which might make the influence of using the CATs undetectable in these scores. Nevertheless, the slight change of students' scores in the *High Insight* group indicated initially some positive influence, although it disappeared later on. It seems the more insights teachers gained from using CATs, the more their students' mathematics scores tended to increase – albeit the differences were very small. Another positive finding is that despite teachers spending less time on regular teaching by including the CATs and students thus also having had less time to practice solving the problems, their scores did not decrease on the immediate posttest. The fact that taking some time to carry out CATs does not really lead students to underperform may encourage Chinese teachers to go beyond straightforward assessment of the results of learning in terms of basic knowledge and skills as included in the textbook (Liu, 2012).

In sum, our study provides evidence that Chinese primary mathematics teachers *can* understand the CATs and integrate them into their practice to gain insights into their students' understanding of mathematics. However, for the majority of the teachers, it seems to be necessary to reserve more time and provide more support, than they got in our study, before they can fully understand the potential of the CATs and benefit from using the CATs to enhance students' learning. Using the CATs implies a strong formative approach to assessment, which for Chinese primary teachers, who often put more emphasis on providing information to students than on getting information about students (Cai & Wang, 2010), may mean a change of perspective. Our CATs can help teachers to develop a more formative approach to assess students' learning, especially in the sense that CATs provide teachers with concrete assessment activities that they can immediately put into classroom practice. But certainly more research is necessary, especially studies that investigate how teachers' culturally-based beliefs about teaching affects their formative use of assessment and that examine how to support teachers to become independent users of formative assessment.

Acknowledgements

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. All opinions are those of the authors and do not necessarily represent the views of the CSC or NWO. The authors thank Prof. dr. Lianhua Ning in Nanjing Normal University, China, for helping to contact schools, and all the teachers involved in this study 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). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429–458.

Cai, J., Ding, M., & Wang, T. (2014). How do exemplary Chinese and U.S. mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*, 265–280.

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.

De Lange, J. (1999). *Framework for classroom assessment in mathematics*. Madison, WI: NICLA/WCER.

Gallego-Arrufat, M.J. & Dandis, M. (2014). Rubrics in a secondary mathematics class. *Mathematics Education, 9*(1), 75–84. Retrieved from http://www.mathedujournal.com/dosyalar/IJEM_v9n1_6.pdf

Gu, F., & Gu, L. (2016). Characterizing mathematics teaching research mentoring in the context of Chinese lesson study. *ZDM Mathematics Education, 48*, 441–454.

Harlen, W. (2007). *Assessment of learning*. London: Sage.

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.

James, M., & McCormick, R. (2009). Teachers learning how to learn. *Teaching and Teacher Education, 25*, 973–982.

Jiangsu Phoenix Education Publishing House. (2014). *Sujiaoban jiaokeshu (xiaoxue shuxue sannianji xiace)* [Sujiaoban Textbook (Mathematics textbook for Grade 3 in primary education, Volume 2)]. Nanjing: Author.

Jin, H., & Wong, K. Y. (2015). Mapping conceptual understanding of algebraic concepts: an exploratory investigation involving Grade 8 Chinese students. *International Journal of Science and Mathematics Education, 13*(3), 683–703.

Keeley, P., & Tobey, C. R. (2011). *Mathematics formative assessment: 75 practical strategies for linking assessment, instruction, and learning*. Thousand Oaks, CA: Corwin.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute-by minute and day by day. *Educational Leadership, 63*(3), 18–24.

Li, Y. (2000). Purpose, function and types of mathematics assessment in China: A historical perspective. *ZDM Mathematics Education, 2000*, 192–200.

Li, Y., Chen, X., & Kulm, G. (2009). Mathematics teachers' practices and thinking in lesson plan development: A case of teaching fraction division. *ZDM Mathematics Education, 41*, 717–731.

Lin, P. J. (2006). Conceptualization of teachers' understanding of students' mathematical learning by using assessment tasks. *International Journal of Science and Mathematics Education, 4*(3), 545–580.

Liu, J. (2012) *Xiaoxue dinianji shuxue jiaokeshu zhong de xiti yanjiu – jiyu zhongguo xinjiapo jiaokeshu bijiao de shijiao* [Study in mathematics textbook tasks for lower grades in primary school – From the perspective of comparison between textbooks in China and that in Singapore] (Master's thesis). Retrieved from the China National Knowledge Infrastructure.

Ministry of Education of the People's Republic of China (MoE) (2001a). *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) (2001b). *Quanrizhi yiwu jiaoyu shuxue kecheng biaozhun (shiyangao)* [Mathematics curriculum standards of nine-year compulsory education (trial version)]. Beijing, China: Beijing Normal University Press.

Ministry of Education of the People's Republic of China (MoE) (2011). *Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

Panizzon, D., & Pegg, J. (2007). Assessment practices: Empowering mathematics and science teachers in rural secondary schools to enhance student learning. *International Journal of Science and Mathematics Education, 6*(2), 417–436.

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.

Phelan, J., Choi, K., Niemi, D. N., Vendlinski, T., 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.

Phelan, J., Kang, T., Niemi, D. N., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessments in middle school mathematics*. CRESST Report 750. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=869D1839672D5CCB18295951BDF08CE4?doi=10.1.1.175.2357&rep=rep1&type=pdf

Shavelson, R.J. (1973). *The basic teaching skill: Decision making* (R & D Memorandum No.104). Stanford, CA: Stanford University, School of Education, Center for R & D in Teaching.

Schoenfeld, A. H. (2015). Summative and formative assessments in mathematics supporting the goals of the Common Core Standards. *Theory Into Practice, 54*(3), 183–194.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

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.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11–18.

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014a). 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. (2014b). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE, 9*(1), e86817.

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2017). *Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement*. Manuscript submitted for publication.

Wiliam, 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.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice, 22*(1), 140–160.

Zhang, D., & Lee, P. Y. (1991). Examination culture and mathematics teaching. *Proceedings of ICMI – China Regional Conference of Mathematical Education at Beijing, 1*, 8.

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. (2017). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education*, Advance online publication.

Zhu, Y., Yu, W., & Cai, J. (2018) Understanding students' mathematical thinking for effective teaching: A comparison between expert and nonexpert Chinese elementary mathematics teachers. *EURASIA Journal of Mathematics, Science and Technology Education. 14*(1), 213–224.

Authors' contributions

This paper was a collaborative work of the three authors. XZ, MHP, and MV all participated in designing the study. XZ carried out the data collection in Nanjing, China, and was responsible for interpreting the Chinese resources. XZ, MHP, and MV analyzed the data, and drafted and revised the manuscript. All authors read and approved the final manuscript.

Chapter 6

Summary and discussion

Summary and discussion

1. The Improving Classroom Assessment in China project

Since the very beginning of this century, the assessment reform towards using assessment to support teaching and learning has been carried out in mainland China (MoE, 2001). After more than a decade, few studies have investigated Chinese primary school mathematics teachers' assessment at classroom level. Moreover, clear calls have been made, by both the Chinese academic community and teachers themselves, for more help to enhance teachers' assessment practice. With this background, the Improving Classroom Assessment in China (ICA-C) project, as an extension of the Improving Classroom Assessment (ICA) project in the Netherlands, was set up in 2012.

The first goal of the ICA-C project was to update the knowledge about the current situation of Chinese primary school mathematics teachers' perception and practice of classroom assessment. *Classroom assessment*, or formative assessment in the hands of teachers, is the assessment that teachers continuously do during teaching with the aim of collecting information about students' learning and make adequate instructional decisions to meet students' needs (Veldhuis & Van den Heuvel-Panhuizen, 2014a, 2014b, 2017). To investigate how Chinese primary school mathematics teachers consider and perform classroom assessment, two studies were set up. In the first study, we analyzed teacher-written journal papers addressing classroom assessment, and in the second study, we collected new data via a questionnaire on teachers' perception and practice of classroom assessment.

The second goal of the ICA-C project was to explore the possibility for improving Chinese primary school mathematics teachers' formative 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 (Veldhuis & Van den Heuvel-Panhuizen, 2014b, 2017). Two further studies were set up to look into how Chinese primary school mathematics teachers use CATs and examine what new insights they can gain into their students' mathematical understanding from using CATs. In sum, four studies were carried out in the ICA-C project and are reported on in this PhD thesis. The main research findings of each study are summarized as follows.

1.1 Classroom assessment in the eyes of Chinese primary school mathematics teachers

To shed light on teachers' perceptions of the current situation of classroom assessment in primary school mathematics education in China, a review study was conducted based on papers written by teachers in which their experience of classroom assessment and tips learned from others were described and discussed (*Chapter 2*). Through a number of steps of paper selection, 266 teacher-written papers addressing classroom assessment in primary school mathematics education published in 2011 and 2012 were identified in the China National Knowledge Infrastructure database. For analyzing 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 Chinese mathematics curriculum standards (MoE, 2011). This included the purpose of assessment, the content of 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%) at least a quarter of the text referred to feedback, in 64 of these, the teacher-authors considered classroom assessment to be equivalent to feedback. However, it was in only nine papers (3%) that the teacher-authors explained an instructional adaptation based on the gathered assessment information. Regarding the relation between teachers' conceptions of classroom assessment, as revealed in the teacher-written papers and the assessment guidelines in the mathematics curriculum standards, nearly all aspects echoed quite well, except for using assessment results for instructional adaptation. In 142 papers (53%) the teacher-authors either explicitly stated that they conducted assessment using the official assessment guidelines, or paraphrased or cited these guidelines.

1.2 Chinese primary school mathematics teachers' assessment profiles

To sketch a more comprehensive picture of Chinese primary school mathematics teachers' views on assessment, a large-scale survey was carried out (*Chapter 3*). For data collection, an existing questionnaire, previously developed in the Netherlands for investigating teachers' assessment practice and beliefs (Veldhuis & Van den Heuvel-Panhuizen, 2014a; Veldhuis, Van den Heuvel-Panhuizen, Vermeulen, & Eggen, 2013), was adjusted to fit to 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 analyzed. 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 to use several assessment methods for different purposes of instructional decision-making with an average frequency. To assess students, they reported using a number of different problem formats. Also, 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. A second group of teachers belonged to the *Enthusiastic assessors* profile (21.7%), who had above average scores overall. They reported to use 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 important or useful.

1.3 Chinese primary school mathematics teachers' use of CATs

After having acquired some understanding of the current situation of classroom assessment in mathematics in primary schools, the focus of research switched to how teachers' classroom assessment practice could be improved. In an explorative study, Chinese primary school mathematics teachers' use of CATs was investigated (*Chapter 4*). Six female third-grade mathematics teachers and their 216 students from two primary schools in Nanjing, China, participated in this study. Based on the textbook these teachers used, a series of CATs were designed to assess their students' understanding of the division of three-digit numbers by a one-digit number. 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. Also, the teachers were informed to freely adapt the CATs to their own situation, rather than carrying them out as following a fixed recipe. During the two weeks of division teaching, four one-hour meetings were organized to support the teachers, 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, lesson observation and student work were used, if necessary, to mirror or supplement the information reported by the teachers.

According to the teachers' responses, it was found that, although CATs were quite new to them, they could easily include CATs in their practice by changing them to fit to their pre-arranged lesson plans. These changes included organizational changes such as reducing the number of tasks in CATs, changing time for carrying out CATs, or directly teaching students to solve the tasks before or during the assessment. By using the CATs, the teachers also 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 of 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. Nevertheless, they used the information to give their students, during or after carrying out the CATs, instant help for arriving at correct answers. Also, the teachers used the teacher guide of using CATs to adapt their instruction beforehand. In this way, the CATs, instead of being implemented as assessment activities, were often included as extra exercises in the teachers' pre-arranged lesson plans. Overall, the teachers found that the CATs were feasible to conduct, revealing students' learning information, efficient when collecting the information, and also helpful to engage their students in assessment.

1.4 Insights Chinese primary school mathematics teachers gained from using CATs

In this intervention study, it was examined what new insights Chinese primary school mathematics teachers gained into their students' mathematical understanding from using CATs (*Chapter 5*). The setup of this study was similar to that in the explorative study (*Chapter 4*); however, a larger group of teachers was involved, and a different mathematics topic was chosen. To be more specific, 25 third-grade mathematics teachers from nine schools in Nanjing were provided with eight CATs assessing students' understanding of multiplication of two-digit numbers. 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 new insights from the CATs. When teachers referred to the mathematics 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 mathematics tests were analyzed 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 less 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 favored 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 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, at least two-third of the students got more than 85 (out of 100) points. After the intervention, there was a very small positive change in the scores for the students in the *High Insight* group, and a very small negative change in the *Some Insight* and *No Insight* group. However, when performing an analysis of covariance by putting standardized pretest score as covariate and insight group as fixed factor, no significant effect of having teachers' gained insights on student achievement was found.

2. Practical implications

The findings of the four studies in the ICA-C project point out several directions in which different stakeholders, like policy makers, educational researchers, teacher educators, and teachers can make efforts in order to improve classroom assessment in mathematics in primary school education.

The most unexpected finding across different studies in the ICA-C project is that using assessment information to adapt further teaching seems to be a missing piece in Chinese primary school mathematics teachers' picture and practice of classroom assessment. Although the idea of using assessment to improve teaching is overall embraced by teachers, relevant practice is hardly mentioned in the teacher-written papers (*Chapter 2*) and no evidence was found that the teachers used the student learning information gained from CATs to adjust their further teaching (*Chapter 4*). In fact, many teachers reported assessment did not influence their teaching (*Chapter 3*). Classroom assessment can only function formatively when the collected information is actually used by teachers to adapt 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 Chinese teachers do not really use assessment formatively, as their teaching is not shaped in light of the evidence of students' learning elicited by assessment. This may have to do with Chinese mathematics teachers' established conception of effective teaching. More emphasis is put on teachers' skills in providing information to students than the skill of gathering information about students (Cai & Wang, 2010). Another possible reason concerns the professional support Chinese primary school mathematics teachers have received. By analyzing a number of post-lesson debriefs given by teaching research specialists, Gu and Gu (2016) found little time (2% of the total mentoring time) was spent on formative assessment of students learning. Moreover, the suggestions about improving further teaching were mainly based on teaching research specialists' past experiences rather than the evidence of student learning emerged in class. Therefore, they concluded, "[t]his type of mentoring may lead teachers to focus on their teaching strategies and skills rather than soliciting students thinking and learning as recommended in the new curriculum" (Gu & Gu, 2016, p. 451). This implies that maybe not only teachers, but also those who support teachers to implement assessment formatively, lack the awareness of using assessment information to adapt further teaching. Therefore, it is suggested that stakeholders at district and province level in the Chinese educational system pay more attention to classroom assessment.

Nevertheless, for other key aspects of assessment explained in the assessment guidelines in Chinese mathematics curriculum standards (MoE, 2011), including the purpose, the content, the assessor, the methods of assessment, and the provision of feedback, teachers' conceptions echoed quite well what is promoted by the Ministry of Education. Notwithstanding this positive finding, the structured presentation of the Chinese assessment guidelines, addressing the key aspects of assessment one-by-one, may have a disadvantage. By providing the assessment guidance in this way, it might not be clear for teachers how these aspects of classroom assessment function as a whole and it remains unclear how to embed them in their daily teaching practice. After all, knowing that teachers need to assess different mathematics content and need to use a variety of assessment methods is one thing; deciding which building blocks are to be assessed by employing which assessment method in a lesson of multiplication of two-digit numbers is another thing. Possible improvements regarding the presentation of the assessment guidelines could be, for example, offering more detailed suggestions and concrete examples illustrating how the key aspects work as a whole and how to integrate assessment into teachers' practice.

Another encouraging finding from the ICA-C project has to do with the potential of using a formative approach to assessment to improve Chinese primary school mathematics teachers' classroom assessment practice. It was found that CATs, as a domain-specific operationalization of such a formative assessment approach, were feasible for teachers to carry out (*Chapter 4*). More importantly, many teachers reported to gain new insights into their students' understanding from using CATs (*Chapter 5*). Although the design of the CATs was based on the Chinese textbook, it was also largely influenced by the principles of RME-based assessment. The results from this research suggest the advantage of opening teachers' eyes regarding assessment by offering questions that originate from a different educational background (Callingham, 2008). In addition, providing teachers with techniques for formative assessment in fact created an opportunity to look into how they conduct assessment in their daily mathematics teaching. Despite the fact that the Chinese teachers were positive about using CATs to bring them new information about student learning and new knowledge about what and how to assess, they mostly first explained to the students how to find the correct answers to the exercises in the CATs. In this sense, using assessment formatively appears to be a real challenge for Chinese primary school mathematics teachers. To facilitate teachers to conduct formative assessment, it would be helpful to provide them with some feasible techniques as a starting point (Lee & Wiliam, 2005). By trying out these new techniques, teachers may demonstrate changes in their assessment activities. The fundamental shift in their beliefs about formative assessment needs to occur before they can make adaptive instructional decisions in light of assessment (Earl, Volante, & Katz, 2011). In sum, concrete and feasible examples about formative assessment activities and support for strengthening teachers' ideas about formative assessment need to go hand in hand.

3. Suggestions for further research

In the ICA-C project, two studies were set up to investigate the current situation of classroom assessment in Chinese primary school mathematics education. Because the educational situation can be very different between regions in China (Wang, Yuan, Tian, & Zhang, 2013), a broader scope was chosen by identifying teacher-written papers in a national database (*Chapter 2*) and distributing the questionnaire in as many provinces as possible (*Chapter 3*). Consequently, teachers' perception and practice of classroom assessment were mainly examined based on teachers' self-reported data. In another two studies (*Chapter 4*, *Chapter 5*) in the ICA-C project aimed at exploring teachers' use of CATs. In the pilot study and the intervention study, only teachers in one city (Nanjing) who used a specific textbook series (Sujiaoban) were involved. To get a comprehensive and thorough understanding of how CATs are used by Chinese teachers and whether CATs can help them to improve their assessment practice, it is necessary to recruit teachers from other regions in China and using different textbooks. Furthermore, only a preliminary attempt was made to explore the relationship between teachers' use of CATs and their students' mathematics achievement (*Chapter 5*). Further research is recommended in which an experimental set up and longer intervention period are used and in which teachers' assessment practice is systematically observed.

Moreover, according to the research findings, distinct groups of Chinese primary school mathematics teachers could be identified depending on their views on assessment (*Chapter 3*) and the insights into students' learning gained from using CATs (*Chapter 5*). It would be interesting to investigate the factors that possibly make *Enthusiastic Assessors* also *High Insight* teachers, and *Unenthusiastic Assessors* also *No Insight* teachers. For example, recently it was revealed significant differences existed between expert Chinese primary school mathematics teachers and their non-expert colleagues in their perception and reported behavior of understanding students' mathematics thinking (Zhu, Yu, & Cai, 2018). Another question raised is whether teachers with different assessment profiles would benefit differently from carrying out CATs or formative assessment in general.

Although the results showed that CATs were feasible and helpful for Chinese primary school mathematics teachers to improve their classroom assessment activities, it is also evident that the teachers faced a real challenge when conducting assessment in a formative way. Several aspects of the teachers' regular teaching activities seemed to hinder them from making the most of using formative assessment, for example, having a full and fixed teaching plan, giving priority to finishing what is in the textbook, and focusing more on the accuracy of answers than on the strategies used by students. Hence, further research that investigates the difficulties Chinese teachers face when using formative assessment in practice needs to be undertaken.

Another issue that needs further deliberation is the role external examination plays in Chinese teachers' classroom assessment activities. Although more and more emphasis has been put on teachers' formative use of assessment since 2001, the selection function of assessment still exerts great influence on teachers' daily teaching. External examination is not only considered as a legitimate source of extrinsic motivation for students' learning, but also to be able to pass examinations has acquired the position of something of value in itself (Leung, 2008). Chinese teachers in general consider the selection function of assessment unified with educational improvement (Brown, Hui, Yu, & Kennedy, 2011). Therefore, another interesting direction for further research would be to investigate how external examination affects teachers' use of classroom assessment.

4. Conclusion

The ICA-C project reported on in this PhD thesis was aimed at investigating the current situation of Chinese primary school mathematics teachers' perceptions and practices of classroom assessment, and exploring the possibility for improving teachers' assessment activities by using CATs. The results show Chinese primary school mathematics teachers in general endorsed the idea of using assessment to improve teaching and learning. They reported to use various methods to assess different types of students' knowledge and skills for a range of purposes, which echoes well to what is advocated in the mathematics curriculum standards in mainland China. Despite these encouraging findings, teachers did not really use the assessment information to adapt their further instruction. In addition, quite some teachers hold a relatively negative view regarding their use of assessment. In terms of improving teachers' classroom assessment, the conclusion was drawn that CATs are feasible and useful for Chinese primary school mathematics teachers to enhance their classroom assessment activities. Nevertheless, this research also suggests that using CATs, which implies a strong approach to formative assessment, is a challenge for Chinese teachers and more support is necessary for helping them make the most of using formative assessment.

References

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.

Black, P., & Wiliam, D. (1998). Inside the black box. Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

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.

Callingham, R. (2008). Perspectives gained from different assessment tasks on Chinese and Australian school students learning mathematics. *Evaluation & Research in Education, 21*(3), 175–187.

Earl, L. M., Volante, L., & Katz, S. (2011). Unleashing the promise of assessment for learning. *Education Canada, 51*.

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.

Gu, F., & Gu, L. (2016). Characterizing mathematics teaching research mentoring in the context of Chinese lesson study. *ZDM Mathematics Education, 48*, 441–454.

Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development, 9*, 265–283.

Leung, F. K. S. (2008). In the books there are golden houses: mathematics assessment in East Asia. *The International Journal on Mathematics Education, 40*(6), 983–992.

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) (2011a). *Yiwu jiaoyu shuxue kecheng biaozhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncct.moe.edu.cn/2014/curriculumPolicy_1115/3175.html

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. (2017). *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.

Wang, S., Yuan, L., Tian, Z., & Zhang, X. (2013). Woguo ge shengfen jiaoyu fazhan shuiping bijiao fenxi [The Comparative Research on the level of regional development of education]. *Research in Education, 6*, 29–41.

Zhu, Y., Yu, W., & Cai, J. (2018) Understanding students' mathematical thinking for effective teaching: A comparison between expert and nonexpert Chinese elementary mathematics teachers. *EURASIA Journal of Mathematics, Science and Technology Education*. *14*(1), 213–224.

Appendices

Appendix A

Papers included in the review (Chapter 2)

#	Author	Year	Title	Journal
1	丁丹 [Ding, D.]	2011	真实的课堂你在我眼中最美 [Real class is the most beautiful class in my eyes]	湖南教育(下) [Education in Hunan]
2	丁琴 [Ding, Q.]	2011	课堂评价不可厚此薄彼 [Classroom assessment should not favor one more than another]	湖北教育(教育教学) [Education in Hubei (Education & Teaching)]
3	严国东 [Yan, G.]	2011	小学数学作业有效性设计之我见 [My opinions on designing valid assignments in primary mathematics education]	新课程(中旬) [New curriculum]
4	何冠 [He, G.]	2011	浅议小学数学作业评价的有效性策略 [Discussion about the strategies of effectively assessing students' assignment]	小学教学参考 [Reference of primary school teaching]
5	何益明 [He, Y.]	2011	小学数学课堂评价的不良倾向及对策 [Undesirable tendency and corrections in classroom assessment in primary mathematics education]	新课程导学 [Guidance of learning new curriculum]
6	何苗, 方彩容 [He, M., & Fang, C.]	2011	浅谈小学数学作业批改 [Some discussion about the correction of students' assignment]	新课程(小学) [New curriculum (primary education)]
7	俞玉花 [Yu, Y.]	2011	依据教材特点凸显评价本质 [Outline the essence of assessment based on the characteristics of textbooks]	新课程(小学) [New curriculum (primary education)]
8	傅志雄 [Fu, Z.]	2011	注重数学课堂评价提高课堂教学效率 [Emphasizing mathematics classroom assessment and improving classroom teaching efficiency]	教育教学论坛 [Forum of education and teaching]
9	傅清清 [Fu, Q.]	2011	艺术性地实施小学数学课堂教学的即时评价 [Implement immediate assessment artistically in primary mathematics teaching]	小学时代(教育研究) [Primary times (education research)]
10	刘丽萍 [Liu, L.]	2011	数学教学评价 – 课堂教学中最艳丽的花朵 [Mathematics teaching and learning assessment: the most gorgeous flower in classroom teaching]	教育革新 [Educational innovation]
11	刘倩 [Liu, Q.]	2011	小学数学课堂评价之我见 [My opinions on classroom assessment in primary mathematics education]	学生之友(小学版)(下) [The friends of students (primary education)]
12	刘国涛 [Liu, G.]	2011	课堂教学评价巧妙设想 [Ideas and imagination of classroom teaching and learning assessment]	陕西教育(行政) [Education in Shaanxi (administration)]
13	刘平国 [Liu, P.]	2011	小学数学教学与多元评价 [Primary school mathematics teaching and multi-actor assessment]	教育艺术 [Art of education]

#	Author	Year	Title	Journal
14	刘霞	2011	浅谈小学数学课堂提问策略	中国科教创新导刊
	[Liu, X.]		[Some discussion about the strategies of questioning in mathematics classroom in primary school]	[Guidance journal of innovation in technology and education]
15	单亚蓓	2011	追求有效的数学课堂	陕西教育(教学版)
	[Shan, Y.]		[Pursue effective mathematics classroom]	[Education in Shaanxi (teaching)]
16	叶汗青	2011	小学数学“知识与技能”多元评价的探索	小学教学参考
	[Ye, H.]		[Exploration of multi-actor assessment of “knowledge and skills” in primary mathematics education]	[Reference of primary school teaching]
17	吕宁宁	2011	浅谈数学教学中应有的专业性评价	新课程(小学)
	[Lyu, N.]		[Some discussion about assessing professionally in mathematics teaching and learning]	[New curriculum (primary education)]
18	吴丽凤	2011	小学数学作业设计应体现“五性”	陕西教育(行政)
	[Wu, L.]		[“Five principles” in designing mathematics assignment in primary education]	[Education in Shaanxi (administration)]
19	吴卫忠	2011	让数学课堂成为“动感地带”	小学教学参考
	[Wu, W.]		[Turn the mathematics classroom into a M-zone]	[Reference of primary school teaching]
20	吴大明	2011	评语,让批改数学作业活起来	中小学数学(小学版)
	[Wu, D.]		[Written comments make the assessment of mathematics assignment lively]	[Mathematics in primary and second education (primary level)]
21	吴桂柳	2011	小学低年级数学课堂教学评价之我见	小学时代(教育研究)
	[Wu, G.]		[My opinions on mathematics classroom teaching and learning assessment in lower grades in primary education]	[Primary times (education research)]
22	周亮亮	2011	在数学作业中巧用评语	学生之友(小学版)
	[Zhou, L.]		[Use written feedback skillfully when assessing mathematics assignment]	[The friends of students (primary education)]
23	周仁军	2011	浅谈对学生进行数学学习过程的评价	数学大世界(教师适用)
	[Zhou, R.]		[Some discussions about assessment of students during the process of their mathematics learning]	[The world of mathematics (for teachers)]
24	周桂敏	2011	让数学课堂充满活力	小学时代(教育研究)
	[Zhou, G.]		[Make the mathematics classroom full of energy]	[Primary times (education research)]
25	夏小玲	2011	多元评价语言的应用及其价值	教师博览(科研版)
	[Xia, X.]		[The application and value of assessment language given by multi-actors]	[Expo for teachers (research)]
26	夏杰	2011	与孩子一起享受惬意的课堂 – 浅谈小学低年级课堂教学评价语言的策略	阅读与鉴赏(中旬)
	[Xia, J.]		[Enjoy the pleasant classroom with together students: some discussion about strategies of using classroom teaching and learning assessment language in lower grades in primary schools]	[Reading and appreciation]

#	Author	Year	Title	Journal
27	夏翠莲 [Xia, C.]	2011	关注课堂评价提高教学效益 – “三位数加法(不进位加)”案例分析与反思 [Focus on classroom assessment and improve the effectiveness: analysis and reflection on teaching multiplication of three-digit numbers (without carrying)]	小学教学参考 [Reference of primary school teaching]
28	姜平 [Jiang, P.]	2011	对小学数学作业评价的探讨 [Discussion about assessing students’ assignment in primary mathematics education]	青春岁月 [Youthful age]
29	孙金会 [Sun, H.]	2011	数学作业评价方式的探索 [Exploration of the approaches to assessing mathematics assignment]	新课程(中旬) [New curriculum]
30	季明 [Ji, M.]	2011	赏识性语言在小学数学教学中的运用 [Application of positive feedback in primary school mathematics teaching]	新课程学习(下) [Learning of new curriculum]
31	宋文敏 [Song, W.]	2011	创新量化评价凸显人文关怀–谈教学评价中的量化 [Create quantitative assessment to emphasize humanity: discussion about quantification in assessment of teaching and learning]	小学教学设计 [Teaching design in primary education]
32	常恒 [Chang, H.]	2011	对小学数学课堂教学评价的探究 [Exploration of classroom teaching and learning assessment in primary school mathematics education]	教育教学论坛 [Forum of education and teaching]
33	张丽媛 [Zhang, L.]	2011	“三位一体”的数学作业评价 [“Triune” mathematics assignment assessment]	新课程(教研) [New curriculum (teaching and research)]
34	张丽颖, 朱立增 [Zhang, L., & Zhu, L.]	2011	让“评价”激活孩子的心灵 [Use “assessment” to activate students’ minds]	新课程(小学) [New curriculum (primary education)]
35	张扬 [Zhang, Y.]	2011	打造多元评价的数学课堂 [Establish a multi-actor system of assessment in mathematics education]	考试周刊 [Weekly examination]
36	张海霞 [Zhang, H.]	2011	小学数学课堂评价的科学性与合理性 [Scientificity and rationality of using classroom assessment in primary school mathematics education]	小学时代(教育研究) [Primary times (education research)]
37	张秀珍 [Zhang, X.]	2011	营造教学环境构建高效课堂 [Create teaching environment and establish effective classroom]	中国科教创新导刊 [Guidance journal of innovation in technology and education]
38	张静 [Zhang, J.]	2011	小学数学课堂评价的案例分析与思考 [Case study and reflection on classroom assessment in primary school mathematics education]	江苏教育研究 [Educational research in Jiangsu]
39	徐峥 [Xu, Z.]	2011	浅议小学数学教学中的延时评价 [Some discussion about delayed assessment in mathematics teaching in primary school]	新课程研究(上旬刊) [Research in new curriculum]

#	Author	Year	Title	Journal
40	徐敏	2011	即时评价 – 精彩课堂的“导航仪”	小学教学参考
	[Xu, M.]		[The “GPS” of a wonderful classroom: immediate assessment]	[Reference of primary school teaching]
41	徐敏智	2011	数学作业的评价初探	安徽教育
	[Xu, M.]		[Exploration of assessing mathematics assignment]	[Education in Anhui]
42	徐树林	2011	小学数学作业评价之我见	新课程(教研)
	[Xu, S.]		[My opinions on assessing mathematics assignment in primary school]	[New curriculum (teaching and research)]
43	徐满珍	2011	构建师生情感交流平台 – 评语式评价在小学数学作业批改中的思考	现代阅读(教育版)
	[Xu, M.]		[Establish the platform for emotion exchanges between the teacher and students: reflection on written feedback when assessing mathematics assignment in primary school]	[Modern reading (education)]
44	戴小军	2011	让课堂评价成为师生心灵的对话	新课程学习(上)
	[Dai, X.]		[Turn classroom assessment into heart to heart dialogue between the teacher and students]	[Learning of new curriculum]
45	曹志国	2011	数学学习中的四个“学会”	数学大世界(教师适用)
	[Cao, Z.]		[Four “things to be learned” in mathematics learning]	[The world of mathematics (for teachers)]
46	曹永红	2011	试议小学生数学学习评价方式	内蒙古教育
	[Cao, Y.]		[Discussion on the approach to assessing mathematics learning of primary students]	[Education in Inner Mongolia]
47	朱晓杰	2011	浅谈小学数学教学中的多元评价	中国科教创新导刊
	[Zhu, X.]		[Some discussion on the multi-actor assessment in primary school mathematics teaching]	[Guidance journal of innovation in technology and education]
48	权克华	2011	小学数学作业发展性评价探微	新课程学习(上)
	[Quan, K.]		[Exploration of developmental assessment of mathematics assignment in primary schools]	[Learning of new curriculum]
49	李云香	2011	正确使用评价手段提高学生数学素养	成才之路
	[Li, Y.]		[Use assessment correctly to improve students’ mathematics literacy]	[The road of being a useful person]
50	李占文	2011	对小学数学课堂教学激励性评价的反思	学周刊
	[Li, Z.]		[Reflection on motivating assessment in primary school mathematics classroom teaching]	[Weekly learning]
51	李志新	2011	如何让数学课堂教学评价恰到好处	学生之友(小学版)
	[Li, Z.]		[How to make mathematics classroom teaching and learning assessment to the point]	[The friends of students (primary education)]
52	李忠玉	2011	农村小学数学课堂教学的有效性及思考	新课程学习(上)
	[Li, Z.]		[Reflection on the effectiveness of mathematics classroom teaching in rural primary schools]	[Learning of new curriculum]
53	李静	2011	生动有趣的评价个性飞扬的课堂	新课程(小学)
	[Li, J.]		[Interesting assessment and the classroom with personality]	[New curriculum (primary education)]

#	Author	Year	Title	Journal
54	杜浩浪 [Du, H.]	2011	浅谈小学数学中的作业评价 [Some discussion about assignment assessment in primary school mathematics education]	数学大世界(教师适用) [The world of mathematics (for teachers)]
55	杨森 [Yang, S.]	2011	小学低年级学生作业评价改革之我见 [My opinions on the reform of assignment assessment in lower grades in primary schools]	新课程(教研) [New curriculum (teaching and research)]
56	杨蕾 [Yang, L.]	2011	浅谈小学数学作业的批改 [Some discussion about mathematics assignment assessment in primary schools]	学周刊 [Weekly learning]
57	杨贵芝 [Yang, G.]	2011	数学课堂教师提问艺术 [The art of teachers' questioning in mathematics class]	中国教育技术装备 [China educational technology & equipment]
58	杨项余 [Yang, X.]	2011	再议让作业批改更有效 [Discussion again about making assignment correction more effective]	中小学数学(小学版) [Mathematics in primary and second education (primary level)]
59	柏茂花 [Bai, M.]	2011	让孩子的数学作业本温馨情趣起来 [Make students' mathematics notebooks more interesting]	新课程学习(上) [Learning of new curriculum]
60	武志红 [Wu, Z.]	2011	浅谈数学教学评价 [Some discussion about mathematics teaching and learning assessment]	河北经贸大学学报(综合版) [Academic journal of Hebei University of Economics and Business (comprehensive version)]
61	汤曙红 [Tang, S.]	2011	重构学习评价提升教学品味 [Restructure learning assessment and improve the taste of teaching]	湖南教育(下) [Education in Hunan]
62	沈文胜 [Sheng, W.]	2011	小学数学课堂提问策略 [Strategies of questioning in primary mathematics class]	福建基础教育研究 [Research on basic education in Fujian]
63	王丹 [Wang, D.]	2011	适时评价激发潜能 [Appropriate assessment can stimulate potential]	小学时代(教师) [Primary times (teachers)]
64	王丽 [Wang, L.]	2011	优化过程性评价,让数学课堂焕发生命光彩 [Optimize assessment of the process and make mathematics classroom more lively]	小学教学参考 [Reference of primary school teaching]
65	王从萍 [Wang, C.]	2011	数学课堂呼唤真实有效的评价 [Mathematics classes ask for real and effective assessment]	小学时代(教师) [Primary times (teachers)]
66	王向红 [Wang, X.]	2011	基于新课程的小学数学课堂教学即时评价 [Immediate assessment in primary school mathematics teaching in the background of new curriculum reform]	教育教学论坛 [Forum of education and teaching]
67	王文华 [Wang, W.]	2011	用评价绽放学生的心灵之花 [Use assessment to make the flowers in students' heart blossom]	中国教育技术装备 [China educational technology & equipment]

#	Author	Year	Title	Journal
68	王明清	2011	立足过程促进发展 – 试论数学教学评价的时弊与改进对策	新课程研究(上旬刊)
	[Wang, M.]		[Focus on the process to facilitate improvement: discussion about problems and correction in mathematics teaching and learning assessment]	[Research in new curriculum]
69	王晓艳	2011	课堂评价应关注学生的状态	辽宁教育
	[Wang, X.]		[Classroom assessment should focus on students' situation]	[Education in Liaoning]
70	王江	2011	原来可以如此美丽 – 浅谈小学数学课堂中的评价语言	数学之友
	[Wang, J.]		[It could be so beautiful: some discussion about the assessment language used in primary mathematics classes]	[The friends of mathematics]
71	王班利	2011	浅论有效提问提高课堂效率	新课程(教研)
	[Wang, B.]		[Some discussion about the effectiveness of effective questioning in class]	[New curriculum (teaching and research)]
72	王秋珍	2011	小学数学课堂的语言艺术	快乐阅读
	[Wang, Q.]		[The art of language in primary mathematics classes]	[Reading happily]
73	王红顺	2011	从"旁观"的角度"热议"中小学管理的"新政"(下) – 学校管理见闻"赏析"	教书育人
	[Wang, H.]		["Heated debate" on "new policy" in the management in primary and secondary school from the perspective of an "outsider": "appreciation" about school management]	[Teaching knowledge and educating people]
74	王艳青	2011	如何改进数学作业评价	考试周刊
	[Wang, Y.]		[How to improve mathematics assignment assessment]	[Weekly examination]
75	王颖	2011	让学生在数学课堂评价中快乐学习	齐齐哈尔大学学报(哲学社会科学版)
	[Wang, Y.]		[Make students learn happily in mathematics classroom in primary schools]	[Academic journal of Qiqihar University (Philosophy and social sciences)]
76	白雪莲	2011	谈小学数学思维能力的课堂评价	教育教学论坛
	[Bai, X.]		[Discussion about classroom assessment of students' thinking]	[Forum of education and teaching]
77	祝丽华	2011	有效的作业评价	学周刊
	[Zhu, L.]		[Effective assignment assessment]	[Weekly learning]
78	符美芬	2011	浅论小学数学作业批改方法	考试周刊
	[Fu, M.]		[Some discussion about the ways of correcting mathematics assignment in primary education]	[Weekly examination]
79	管燕萍	2011	课堂评价语——成功教学的法宝	品牌(理论月刊)
	[Guang, Y.]		[Classroom assessment language: the secret of successful teaching]	[Brand (theory monthly)]
80	胡建华	2011	小学数学课堂教学中,教师评价方式的再探索	考试与评价
	[Hu, J.]		[Re-exploration of teachers' assessment in primary school mathematics teaching]	[Examination and assessment]
81	胡静	2011	小学数学过程性评价中主体作用的探析	剑南文学(经典教苑)
	[Hu, J.]		[Exploration and analysis about who is the assessor in process assessment in primary mathematics education]	[Jiannan arts (classical teaching)]

#	Author	Year	Title	Journal
82	葛艳辉	2011	小学数学作业设计与评价的探究	中国教师
	[Ge, Y.]		[Exploration of design and assessment of mathematics assignment in primary education]	[Chinese teachers]
83	董志军	2011	数学课堂需要什么样的表扬	小学教学研究
	[Dong, Z.]		[What kind of praise is needed in mathematics classroom]	[Research on teaching in primary education]
84	董文歆	2011	让评价扬起学生发展的风帆 – 小学数学发展性评价策略的实践研究	小学教学参考
	[Dong, W.]		[Use assessment raise the sailing of students' development: empirical research on process assessment in primary mathematics education]	[Reference of primary school teaching]
85	蒋滢	2011	合理的数学课堂评价是学生学习的催化剂	小学时代(教育研究)
	[Jiang, Y.]		[Reasonable mathematics classroom assessment is the catalyst of students' learning]	[Primary times (education research)]
86	薛栋	2011	浅谈《数学课程标准》中的学生评价	数学学习与研究
	[Xue, D.]		[Some discussion about student assessment in the mathematics curriculum standards]	[Mathematics learning and research]
87	许友良	2011	发挥评价魅力,增强学生信心	小学时代(教师)
	[Xu, Y]		[Exert the influence of assessment on enhancing students' confidence]	[Primary times (teachers)]
88	许红飞	2011	优化数学作业减轻学业负担	新课程(小学)
	[Xu, H.]		[Optimize mathematics assignment and reduce students' learning loads]	[New curriculum (primary education)]
89	赵军	2011	关于数学作业批改方式的实践与思考	学生之友(小学版)(下)
	[Zhao, J]		[Practice and refection on the ways of correcting mathematics assignment]	[The friends of students (primary education)]
90	赵崇雷	2011	真诚的评价让课堂焕发生命的活力	科技致富向导
	[Zhao, C.]		[Use sincere assessment to make the classroom lively]	[Guidance of using technology to make a better life]
91	赵彩文	2011	浅谈即时评价在数学课堂教学中的应用策略	小学教学参考
	[Zhao, C.]		[Some discussion about the strategy of using immediate assessment in mathematics classroom teaching]	[Reference of primary school teaching]
92	赵荣华	2011	"评价与反思"的教学探索	教育研究与评论(小学教育教学)
	[Zhao, R.]		[Educational exploration of "assessment and reflection"]	[Educational research and comment (teaching in primary school)]
93	赵顺天	2011	巧妙评价,促进发展	考试周刊
	[Zhao, S.]		[Skillful assessment to facilitate improvement]	[Weekly examination]
94	郑伟东	2011	新理念下数学课堂教学中的教师评价	小学教学参考
	[Zheng, W.]		[Teacher assessment in mathematics classroom teaching under the background of new curriculum reform]	[Reference of primary school teaching]
95	郑佩兰	2011	新课程下的小学数学课堂评价	学生之友(小学版)
	[Zheng, P.]		[Primary school mathematics classroom assessment under the new curriculum reform]	[The friends of students (primary education)]

#	Author	Year	Title	Journal
96	郭伟 [Guo, W.]	2011	让“无情”的评价变为“有情”的激励 – 刍议新课程理念下小学数学课堂评价的策略 [Turn “unmerciful” assessment into “compassionate” stimulation: discussion about the strategies of primary school mathematics classroom assessment]	小学教学参考 [Reference of primary school teaching]
97	钟凤娇 [Zhong, F.]	2011	浅谈小学数学教学评价 [Some discussion about primary school mathematics classroom assessment]	数学学习与研究 [Mathematics learning and research]
98	钱立琴 [Qian, L.]	2011	把握学生心理提高数学课堂效率 [Keep students’ psychological situation to improve effectiveness of mathematics classroom]	新课程学习(上) [Learning of new curriculum]
99	闫黎明 [Yan, L.]	2011	怎样进行小学数学教学的课堂评价 [How to conduct classroom assessment in mathematics teaching in primary education]	山东教育 [Education in Shandong]
100	陆洪荣, 王志成 [Lu, H., & Wang, Z.]	2011	写好数学日记提高综合能力 – 指导小学生写数学日记初探 [Write mathematics diary to improve ability comprehensively: exploration of guiding primary students’ to write mathematics diary]	科学咨询(教育科研) [Scientific consultation (educational research)]
101	陈严 [Chen, Y.]	2011	“炸”出的一段美丽 – 小学数学课堂评价研究案例 [Turn out to be a beautiful story: case study of classroom assessment in primary education]	数学大世界(教师适用) [The world of mathematics (for teachers)]
102	陈国彩 [Chen, G.]	2011	浅谈数学作业二次评价 [Some discussion about assessing mathematics assignment for the second time]	新课程(小学) [New curriculum (primary education)]
103	陈宝元, 刘贇 [Chen, B., & Liu, Y.]	2011	也谈“作业积分”的魅力 [Discussion of the charm of “accumulative reward system of incentives”]	中小学数学(小学版) [Mathematics in primary and second education (primary level)]
104	陈微 [Chen, W.]	2011	有效评价,让数学课堂灵动起来 [Effective assessment to make mathematics classroom energetic]	考试周刊 [Weekly examination]
105	陈晓慧 [Chen, X.]	2011	小议课堂评价 [Discussion about classroom assessment]	读与写(教育教学刊) [Reading and writing (education and teaching)]
106	陈爱玲 [Chen, A.]	2011	小学数学课堂上的语言评价艺术 [The art of assessment language used in primary mathematics classroom]	教师博览(科研版) [Expo for teachers (research)]
107	陈琼瑶 [Chen, Q.]	2011	有效理答,让数学课堂更精彩 [Effective questioning and answering will make classroom assessment more wonderful]	数学大世界(教师适用) [The world of mathematics (for teachers)]
108	陈秀华, 马小芳 [Chen, X., & Ma, X.]	2011	激励评价——课堂因此而精彩 – 小学数学课堂评价新探 [Stimulative assessment the reason of making classroom wonderful: new exploration of classroom assessment in primary school mathematics education]	现代阅读(教育版) [Modern reading (education)]

#	Author	Year	Title	Journal
109	陈素珍 [Chen, S.]	2011	小学数学课堂上如何发挥评价的作用 [How to exert the influence of assessment in primary mathematics classroom]	新课程研究(上旬刊) [Research in new curriculum]
110	陈花 [Chen, H.]	2011	课堂评价——学生幸福成长的天地 [Classroom assessment: the world of making students develop happily]	数学学习与研究 [Mathematics learning and research]
111	陈露 [Chen, L.]	2011	数学课堂评价语言的艺术 [The art of using classroom assessment language in mathematics education]	教育教学论坛 [Forum of education and teaching]
112	隋颖 [Sui, Y.]	2011	作业评价与学生能力发展研究 [Research on assignment assessment and development of students' ability]	才智 (Wisdom)
113	顾幼芳 [Gu, Y.]	2011	小学数学课堂评价管窥 [Have a view on primary mathematics classroom assessment]	数学学习与研究 [Mathematics learning and research]
114	顾德风 [Gu, D.]	2011	小学数学课堂评价问题分析与对策的研究 [Problems and suggestions of classroom assessment in primary school mathematics education]	考试周刊 [Weekly examination]
115	颜家骐 [Yan, J.]	2011	充分利用课堂评价促进学生可持续发展 [Make use of classroom assessment to facilitate students' sustainable development]	教师 [Teacher]
116	颜清怀 [Yan, Q.]	2011	论新课程背景下小学数学教学的评价策略 [Discussion about the strategies of assessment in primary school mathematics education under the background of new curriculum reform]	考试周刊 [Weekly examination]
117	马瑜 [Ma, Y.]	2011	有效运用引导、指导和评价语言提高小学数学教学有效性 [The effectiveness of guiding and using assessment language to improve primary school mathematics teaching]	教育教学论坛 [Forum of education and teaching]
118	马雪丽 [Ma, X.]	2011	“丰富”数学作业评价功能 [“Enrich” the function of mathematics assignment assessment]	新课程(教研) [New curriculum (teaching and research)]
119	骆丽珊 [Luo, L.]	2011	让作业评价成为学生的“氧气” [Make the assignment assessment to become students' “oxygen”]	吉林教育 [Education in Jilin]
120	高向红 [Gao, X.]	2011	小学数学考试评价改革初探 [First exploration of the reform of examination and assessment in primary school mathematics education]	新课程(教研) [New curriculum (teaching and research)]
121	高明 [Gao, M.]	2011	浅谈新课标下小学数学课堂教学评价的误区及对策 [Some discussion about misunderstanding and correction in classroom teaching and learning assessment in primary mathematics education in new curriculum reform]	学生之友(小学版)(下) [The friends of students (primary education)]
122	魏九飞 [Wei, J.]	2011	正确有效评价学生 [Assess students correctly and effectively]	现代教育科学(小学教师) [Modern educational science (primary school teachers)]

#	Author	Year	Title	Journal
123	黄晓波 [Huang, X]	2011	浅谈一年级数学学困生的有效转化 [Some discussion about helping Grade 1 students with learning difficulties in mathematics]	黑河教育 [Education in Heihe]
124	丁佳丽 [Ding, J.]	2012	别样的批语别样的精彩 [Different written feedback different splendidness]	新课程(小学) [New curriculum (primary education)]
125	丁栋青 [Ding, D.]	2012	打开瓶颈:数学学困生的有效转化 [Break the bottleneck: effectively help students with learning difficulties in mathematics]	小学教学参考 [Reference of primary school teaching]
126	侯春玲 [Hou, C.]	2012	巧妙评价张扬个性 [Assess skillfully to show personality]	新课程学习(上) [Learning of new curriculum]
127	俞万军 [Yu, W.]	2012	丰富作业评价手段促进教学相长 [Enrich the ways of assignment assessment to improve teaching and learning]	小学教学参考 [Reference of primary school teaching]
128	俞芳 [Yu, F.]	2012	教学中如何进行有效的课堂评价 [How to carry out classroom assessment effectively during teaching]	现代教育科学(小学教师) [Modern educational science (primary school teachers)]
129	俞霞云 [Yu, X.]	2012	谈教师口头评价的策略 [Talk about the strategies of teachers' verbal assessment]	快乐阅读 [Reading happily]
130	倪莹芝 [Ni, Y.]	2012	探究新课程激活课堂评价 [Explore new curriculum reform and activate classroom assessment]	新课程研究(上旬刊) [Research in new curriculum]
131	冯卫星 [Feng, W.]	2012	小学数学课堂教学评价的思考与研究 [Reflection on and research in classroom teaching and learning assessment in primary mathematics education]	小学教学参考 [Reference of primary school teaching]
132	冯巧南 [Feng, Q]	2012	以"评"促"进" – 谈谈数学教学中的评价 [Use "assessment" to enhance "improvement": discuss about assessment during mathematics teaching]	考试(综合版) [Examination (comprehensive version)]
133	刘世超 [Liu, S.]	2012	科学有效评价促进全面发展 – 浅谈小学数学如何实施有效评价 [Facilitate overall development in a scientific and effective way: discussion about how to conduct assessment in primary school mathematics education]	教师 [Teacher]
134	刘凤霞 [Liu, F.]	2012	精彩评价语,让数学课堂熠熠生辉 [Wonderful assessment language makes classroom assessment shining]	课程教育研究 [Educational research on curriculum]
135	刘卫群 [Liu, W.]	2012	对数学课堂教学评价方法的探究 [Exploration of the strategies of mathematics classroom teaching and learning assessment]	新课程(小学) [New curriculum (primary education)]
136	刘春 [Liu, C.]	2012	改进数学作业评价的点滴思考 [Some reflection on improving mathematics assignment assessment]	中国校外教育 [Education out of school in China]

#	Author	Year	Title	Journal
137	刘红美	2012	课堂评价艺术让童心飞扬	语数外学习(数学教育)
	[Liu, H.]		[The art of classroom assessment make students' heart flying]	[Learning of Chinese, mathematics, & English (mathematics education)]
138	叶春红	2012	论数学教学的课堂评价语言	数学学习与研究
	[Ye, C.]		[Discuss classroom assessment language in mathematics teaching]	[Mathematics learning and research]
139	叶雪芳	2012	有效评价构建和谐数学课堂	吉林教育
	[Ye, X.]		[Assess effectively to establish harmonious mathematics classroom]	[Education in Jilin]
140	吕科	2012	品味精彩 – 由"名师"课堂评价语言想到的	小学时代(教师)
	[Lyu, K.]		[Taste wonderfulness: what is thought about based on expert teachers' classroom assessment language]	[Primary times (teachers)]
141	吴丽春	2012	在小学数学作业中使用"积分"的探索	内蒙古教育
	[Wu, L.]		[Exploration of using "incentives" in mathematics assignment in primary education]	[Education in Inner Mongolia]
142	吴梅林	2012	魅力课堂需要智慧评价	学园(教育科研)
	[Wu, M.]		[Charming classes need wisdom assessment]	[Academic garden (educational research)]
143	吴海燕	2012	表现性评价让数学实践活动课更加灵动	中国校外教育
	[Wu, H.]		[Performance assessment makes mathematics activities more lively]	[Education out of school in China]
144	吴清华	2012	小学数学作业评价之我见	江西教育
	[Wu, Q.]		[My opinions on mathematics assignment assessment in primary education]	[Education in Jiangxi]
145	吴贞金	2012	数学作业评价要注重"三收获"	学生之友(小学版)(下)
	[Wu, Z.]		["Three gains" when paying more attention to mathematics assignment assessment]	[The friends of students (primary education)]
146	唐仁钧	2012	智慧的评价温暖的课堂 – 华应龙老师数学课堂评价语言例谈	湖南教育(下)
	[Tang, R.]		[Assessment with wisdom warms the classroom: discuss examples of classroom assessment language used by mathematics teacher Yinglong Hua]	[Education in Hunan]
147	姜艳	2012	多措并举,提高小学数学作业的有效性	新课程(小学)
	[Jiang, Y.]		[Use multiple methods to improve the effectiveness of using mathematics assignment]	[New curriculum (primary education)]
148	孙慧娟	2012	试论数学教学的有效性	吉林省教育学院学报(中旬)
	[Sun, H.]		[Try to discuss the effectiveness of mathematics teaching and learning]	[Academic journal of educational institute in Jilin province]
149	孙福建	2012	如何把数学课堂教学评价落到实处	教书育人
	[Sun, F.]		[How to put mathematics classroom teaching and learning assessment into practice]	[Teaching knowledge and educating people]
150	季秀玲	2012	数学作业评价的两种途径	数学大世界(教学导向)
	[Ji, X.]		[Two approaches to assessing mathematics assignment]	[The world of mathematics (teaching guide)]

#	Author	Year	Title	Journal
151	尹霞 [Yin, X.]	2012	数学课堂评价语言也精彩 – 让学生享受教学 [Wonderful mathematics classroom assessment language: let students enjoy being taught]	小学科学(教师论坛) [Science in primary education (forum of teachers)]
152	崔素梅 [Cui, S.]	2012	改革数学作业评价促进学生全面发展 [Reform mathematics assignment assessment to overall improvement of students' development]	考试(教研版) [Examination (educational research)]
153	庞二芳 [Pang, E.]	2012	改变评价方式实行多元评价 – 对新课程数学教学评价操作的思考 [Change assessment methods to establish multi-actor assessment system: reflection on conducting mathematics teaching and learning assessment in new curriculum reform]	启迪与智慧(教育) [Inspiration and wisdom (education)]
154	张书芳 [Zhang, S.]	2012	运用课堂评价促进课堂高效 [Carry out classroom assessment to improve class efficiently]	科学大众(科学教育) [Scientific people (science education)]
155	张兵 [Zhang, B.]	2012	学生素质评价与数学学科相结合的研究与实践 [Research and practice of integration of assessing students' literacy and mathematics]	文理导航(下旬) [Guidance of arts and science]
156	张宇龙 [Zhang, Y.]	2012	浅议评语在小学数学中的运用 [Some discussion about using feedback in primary school mathematics teaching]	现代教育科学(小学教师) [Modern educational science (primary school teachers)]
157	张宝 [Zhang, B.]	2012	新课程标准下的小学数学作业评价 [Mathematics assignment assessment in primary education under the background of the new curriculum standards]	启迪与智慧(教育) [Inspiration and wisdom (education)]
158	张旭明 [Zhang, X.]	2012	小学数学课堂口头评价误区的分析与对策 [Analysis and correction of misunderstanding of using verbal assessment in primary school mathematics education]	新课程(小学) [New curriculum (primary education)]
159	张菊 [Zhang, J.]	2012	农村小学数学综合与实践课程的评价方法 [Assessment methods in mathematics curriculum of comprehension and practice in rural place]	广西教育 [Education in Guangxi]
160	徐美义 [Xu, M.]	2012	优化教师课堂评价丰富学生学习体验 – 例谈小学数学课堂评价用语 [Optimize teachers' classroom assessment to enrich students' learning experience: examples of classroom assessment language used in mathematics class in primary school]	现代阅读(教育版) [Modern reading (education)]
161	徐连进 [Xu, L.]	2012	巧用积极评价,促进学生发展 [Using positive assessment skillfully to improve students' development]	小学时代(教师) [Primary times (teachers)]
162	曹开明 [Cao, K.]	2012	把课堂评价的主动权交给学生 [Give the initiatives of classroom assessment into the hands of students]	江西教育 [Education in Jiangxi]

#	Author	Year	Title	Journal
163	朱佩霞 [Zhu, P.]	2012	谈小学数学教学中过程性课堂教学评价的方法 [Discussion about classroom teaching and learning assessment on process in primary school mathematics education]	考试周刊 [Weekly examination]
164	朱俊 [Zhu, J.]	2012	在数学教学中践行新课程的评价观 [Practice the idea of assessment in new curriculum reform in mathematics teaching]	中学课程辅导(江苏教师) [Assistance of curriculum learning in secondary school (teachers in Jiangsu)]
165	朱全明 [Zhu, Q.]	2012	构建小学数学多元评价体系的实践与思考 [Practice of and reflection on establishing a multi-actor system of assessment in primary school mathematics education]	科学大众(科学教育) [Scientific people (science education)]
166	朱希萍 [Zhu, X.]	2012	以生为本的课堂评价 [Student-centered classroom assessment]	小学教学(数学版) [Teaching in primary school (mathematics)]
167	朱恒东 [Zhu, H.]	2012	有效利用小学数学课堂中的语言评价 [Effectively use verbal assessment in mathematics classroom in primary education]	小学生(教学实践) [Primary school students (teaching practice)]
168	朱艳红 [Zhu, Y.]	2012	浅谈小学数学教学评价 [Some discussion about mathematics teaching and learning assessment in primary education]	学周刊 [Weekly learning]
169	朱茂元 [Zhu, M.]	2012	浅谈小学数学作业评价的策略 [Some discussion about mathematics assignment assessment in primary education]	小学教学研究 [Research on teaching in primary education]
170	朱雪梅 [Zhu, X.]	2012	小学数学作业的有效设计与评价 [Design and assessment of mathematics assignment in primary education]	中国科教创新导刊 [Guidance journal of innovation in technology and education]
171	李安森 [Li, A.]	2012	让有效评价在数学课堂中绽放光彩 [Make effective assessment blossom in mathematics classroom]	学生之友(小学版) [The friends of students (primary education)]
172	李小英 [Li, X.]	2012	浅谈新课程下小学数学课堂教学的即时评价 [Some discussion of immediate assessment in mathematics classroom teaching and learning primary education under the background of new curriculum reform]	新课程(小学) [New curriculum (primary education)]
173	李志胜 [Li, Z.]	2012	注重课堂评价促进学生发展 [Pay attention to classroom assessment to improve students' development]	陕西教育(教学版) [Education in Shaanxi (teaching)]
174	李木宇 [Li, M.]	2012	小学数学课堂评价的几点想法 [Several ideas of primary school mathematics classroom assessment]	数学大世界(教学导向) [The world of mathematics (teaching guide)]
175	李涛 [Li, T.]	2012	换个方式评价学生 [Change to another way to assess students]	新课程学习(上) [Learning of new curriculum]
176	李玉民 [Li, Y.]	2012	新课程下对小学数学练习设计的思考 [Reflection on designing mathematics assignment in primary education in new curriculum reform]	考试周刊 [Weekly examination]

#	Author	Year	Title	Journal
177	李红旗 [Li, H.]	2012	运用即时性评价,构建有效数学课堂 [Use immediate assessment to establish effective mathematics classrooms]	中国校外教育 [Education out of school in China]
178	李美芬 [Li, M.]	2012	立足课堂实效提升作业效度 – 初探“轻负高效”视野下的数学作业优化 [Focus on effectiveness of classroom assignment: exploration of optimizing mathematics assignment under the background of “lower workload but better effectiveness”]	内蒙古教育 [Education in Inner Mongolia]
179	杜俊武 [Du, J.]	2012	新课标下小学数学课堂评价浅议 [Some discussion about mathematics classroom assessment in primary education in the new curriculum reform]	陕西教育(教学版) [Education in Shaanxi (teaching)]
180	杜桂红 [Du, G.]	2012	小学数学试卷命题的人文性与实践性 [Humanity and practice of designing mathematics test paper in primary education]	知识窗(教师版) [Window of knowledge (for teachers)]
181	杨丽萍 [Yang, L.]	2012	小学数学课堂学习评价提要 [Short summary of classroom learning assessment in primary school mathematics education]	小学时代(教育研究) [Primary times (education research)]
182	杨娟 [Yang, J.]	2012	以有效评价促快乐成长 – 对课堂评价有效性的思考 [Promoting students developing happily by assessing effectively: reflection on effectiveness of classroom assessment]	教育教学论坛 [Forum of education and teaching]
183	杨小仁 [Yang, X.]	2012	好“评”带来好“价” – 小学数学课堂评价语言浅析 [Good “assessment” brings good “evaluation”: discussion about classroom assessment language in primary school mathematics education]	知识窗(教师版) [Window of knowledge (for teachers)]
184	杨春金 [Yang, C.]	2012	浅析小学数学学习评价策略 [Analysis of the strategies of learning assessment in mathematics in primary education]	内蒙古教育 [Education in Inner Mongolia]
185	杨晓英 [Yang, X.]	2012	论小学数学在新课改下的教学评价 [Teaching and learning assessment under the assessment reform in primary school mathematics education]	新课程学习(中) [Learning of new curriculum]
186	杨瑜 [Yang, Y.]	2012	数学课堂教学中评价的三重境界 [Three levels of assessment in mathematics classroom teaching]	课程教育研究 [Educational research on curriculum]
187	林丽玲 [Lin, L.]	2012	和谐课堂中的教学评价 [Teaching and learning assessment in harmonious classroom]	黑河教育 [Education in Heihe]
188	林娜 [Lin, N.]	2012	课堂评价语言浅说 [Discussing classroom assessment language]	黑河教育 [Education in Heihe]
189	柴晓超 [Cai, X.]	2012	浅谈如何灵活运用数学课堂评价 [Some discussion about how to flexibly use classroom assessment in mathematics]	学周刊 [Weekly learning]
190	梁静 [Liang, J.]	2012	数学课堂多元评价的探索与尝试 [Exploration of using multi-actor system of assessment in mathematics class]	广西教育 [Education in Guangxi]

#	Author	Year	Title	Journal
191	武习平	2012	浅谈小学数学作业评价	教育教学论坛
	[Wu, X.]		[Some discussion about mathematics assignment assessment in primary school education]	[Forum of education and teaching]
192	武成彬	2012	利用评价培养学生自信心的三点做法	山西教育(教学)
	[Wu, C.]		[Three approaches to using assessment to build students' confidence]	[Education in Shanxi (teaching)]
193	沈冬青	2012	"红花"的背后 – "物质性奖励"盛行现象引起的思考	中小学数学(小学版)
	[Shen, D]		[Behind the "red flowers": reflection on the prevalent phenomenon on using "material incentives"]	[Mathematics in primary and second education (primary level)]
194	沈蕾	2012	构建有效评价体系,提升学生数学素养	考试周刊
	[Shen, L.]		[Establish effective assessment system and improve students' mathematics literacy]	[Weekly examination]
195	游本慧	2012	让学生在多元评价中快乐成长	考试周刊
	[You, B.]		[Make students develop happily in a multi-actor system of assessment]	[Weekly examination]
196	潘建琴	2012	浅谈新课程理念下小学数学作业的多元评价	新课程(小学)
	[Pan, J.]		[Some discussion about multi-actor system of assignment assessment in primary school in new curriculum reform]	[New curriculum (primary education)]
197	焦军显	2012	小学数学新课程下的学生评价	学苑教育
	[Jiao, J.]		[Student assessment in primary school mathematics education under the background of new curriculum reform]	[Academic education]
198	王东平	2012	"延时评价"催生精彩课堂	教育
	[Wang, D.]		["Delayed assessment" makes the classroom wonderful]	[Education]
199	王丽云	2012	浅谈小学数学作业评价	华章
	[Wang, L.]		[Some discussion about mathematics assignment assessment in primary education]	[Magnificent chapter]
200	王丽红	2012	关注教学评价促进学生发展	新课程(下)
	[Wang, L.]		[Focus on teaching and learning assessment and improve students' development]	[New Curriculum]
201	王云	2012	小学低年级数学评价方式的探索	数学大世界(教学导向)
	[Wang, Y.]		[Explore assessment methods in teaching mathematics in lower grades in primary school]	[The world of mathematics (teaching guide)]
202	王婷婷	2012	关注作业反馈促进有效学习	现代中小学教育
	[Wang, T.]		[Focus on feedback to assignment and effectively enhance leaning]	[Modern education in primary and secondary schools]
203	王宏伟	2012	课堂教学的评价艺术	新课程(综合版)
	[Wang, H.]		[The art of assessment of classroom teaching and learning]	[New Curriculum (comprehensive version)]
204	王小兵	2012	评价 – 数学课堂质量的"催化剂"	考试与评价
	[Wang, X.]		[Assessment: "catalyst" of the improvement of the quality of mathematics classroom]	[Examination and assessment]

#	Author	Year	Title	Journal
205	王巧云 [Wang, Q.]	2012	浅谈新课标下数学学习评价的多维性 [Some discussion about multi-dimension of mathematics learning assessment in new curriculum reform]	科学大众(科学教育) [Scientific people (science education)]
206	王海滨 [Wang, H.]	2012	小学数学课堂评价三忌 [Three things to be avoided in classroom assessment in mathematics education in primary schools]	浙江教育科学 [Educational science in Zhejiang]
207	王渊 [Wang, Y.]	2012	数学课堂评价的“三结合”策略 [The strategies of “integration of the three” in mathematics classroom assessment]	课程教材教学研究(小教研究) [Research on curriculum, textbook, & teaching (primary level)]
208	王荣娟 [Wang, R.]	2012	小学数学评价的艺术 [The art of mathematics assessment in primary education]	学周刊 [Weekly learning]
209	王霞 [Wang, X.]	2012	巧妙运用课堂评价语提高教学效果 [Use classroom assessment language skillfully and improve the effect of teaching]	现代阅读(教育版) [Modern reading (education)]
210	王颖 [Wang, Y.]	2012	关注过程评价构建高效课 [Focus assessment on process and establish effective classroom]	小学时代(教师) [Primary times (teachers)]
211	白艺强 [Bai, Y.]	2012	刍议小学数学课堂作业的设计 [Discussion about designing assignment used in primary mathematics classes]	学生之友(小学版)(下) [The friends of students (primary education)]
212	石万锁 [Shi, W.]	2012	谈小学数学课堂即时评价的特点与策略 [Discussion about the characteristics and strategies of immediate assessment used in primary mathematics classes]	学园(教育科研) [Academic garden (educational research)]
213	祁顺成 [Qi, S.]	2012	小学数学预习作业的设计和评价初探 [First exploration of designing and assessing assignment used before class in primary school mathematics education]	青年教师 [Young teachers]
214	祁顺成 黄一青 [Qi, S., & Huang, Y.]	2012	掌握提问技巧提高教学效率 – 高效课堂之小学数学课堂提问策略 [Master the key of questioning and improve the efficiency of teaching: the strategy of questioning in mathematics classes in primary schools]	湖北教育(教育教学) [Education in Hubei (Education & Teaching)]
215	秦勇 [Qin, Y.]	2012	浅谈小学数学课堂评价 [Some discussion about mathematics classroom assessment in primary education]	陕西教育(教学版) [Education in Shaanxi (teaching)]
216	章一星 [Zhang, Y.]	2012	小学数学教学评价的有效性探讨 [Discussion about effectiveness of mathematics classroom teaching and learning assessment in primary education]	新课程(小学) [New curriculum (primary education)]
217	缪玲丽 [Liao, L.]	2012	小学数学课堂教学推进过程中提升交往有效性的策略 [The strategies of promoting the effectiveness of exchange in the process in mathematics classroom teaching in primary schools]	小学时代(教育研究) [Primary times (education research)]

#	Author	Year	Title	Journal
218	罗华 [Luo, H]	2012	浅谈如何提高小学数学课堂教学的有效性 [Some discussion about how to improve the effectiveness of mathematics classroom teaching in primary schools]	萍乡高等专科学校学报 [Academic journal of Pingxiang Junior College]
219	胡宏 [Hu, H.]	2012	寻找开启"试卷讲评"的金钥匙 [Find and use the golden key of "evaluating examination papers"]	小学时代(教师) [Primary times (teachers)]
220	胡清溪 [Hu, Q.]	2012	拨动情感之弦陶冶美好情操 – 浅谈小学数学课堂评价的几点思考 [Dial the strings of emotion and cultivate good sentiment: some reflection on mathematics classroom assessment in primary education]	吉林教育 [Education in Jilin]
221	董相芝 [Dong, X.]	2012	浅谈小学数学作业评价的几种方法 [Some discussion about several strategies of mathematics assignment assessment in primary education]	课程教育研究 [Educational research on curriculum]
222	蒋惠珍 [Jiang, H.]	2012	课堂评价如何从虚假走向真实 [Classroom assessment: how to walk from fake to reality]	教师 [Teacher]
223	蒋银花 [Jiang, Y.]	2012	以人为本开展真评价 – 《圆柱的体积》课堂评价的实践与思考 [Conduct real and people oriented assessment: practice of and reflection on classroom assessment in teaching "volume of a circular cylinder"]	小学教学研究 [Research on teaching in primary education]
224	詹守伟 [Zhan, S.]	2012	让有效的课堂评价成为学生数学学习的导向标 [Make effective classroom assessment become the guidance of students mathematics learning]	小学科学(教师版) [Science in primary education (for teachers)]
225	许娟 [Xu, J.]	2012	谈如何组织有效的小学数学课堂教学反馈 [Discussion about how to effectively organize classroom teaching feedback in primary school mathematics education]	小学教学参考 [Reference of primary school teaching]
226	许晶 [Xu, J.]	2012	数学课堂评价初探 [First exploration of mathematics classroom assessment]	新课程(上) [New Curriculum]
227	许高水 [Xu, G.]	2012	让数学练习、批改及评价有效结合 [Make mathematics practice, correction and evaluation effectively integrated]	云南教育(小学教师) [Education in Yunnan (primary school teacher)]
228	谢海英 [Xie, H]	2012	有效提问智慧理答 – 对小学高年级数学课堂教学的几点思考 [Effective questioning and answering with wisdom: some reflection on mathematics classroom teaching in high grades in primary schools]	小学教学参考 [Reference of primary school teaching]
229	谢碧云 [Xie, B]	2012	数学课堂评价之我见 [My opinions on mathematics classroom assessment]	数学大世界(教学导向) [The world of mathematics (teaching guide)]
230	贾海鸥, 唐云亮 [Jia, H., & Tang, Y.]	2012	教学反馈须有序进行 – 论小学数学教学中的教学反馈 [Instructional feedback should be conducted in order: discussion about instructional feedback during primary school mathematics teaching]	新课程学习(上) [Learning of new curriculum]

#	Author	Year	Title	Journal
231	赵玉辉 [Zhao, Y.]	2012	如何在小学数学教学中进行有效评价 [How to carry out effective assessment during primary school mathematics teaching]	中国校外教育 [Education out of school in China]
232	赵红艳 [Zhao, H.]	2012	新课程下的小学数学教学评价 [Primary school mathematics teaching and learning assessment in new curriculum reform]	学周刊 [Weekly learning]
233	赵静 [Zhao, J.]	2012	数学作业评价多元化 [The multi-actor system of assessment in the assessment of mathematics assignment]	小学时代(教育研究) [Primary times (education research)]
234	邬晓莉 [Wu, X.]	2012	关注学生数学作业的错误 [Focus on mistakes that students make in mathematics assignment]	江西教育 [Education in Jiangxi]
235	邱素娟 [Qiu, S.]	2012	浅谈小学数学作业评价 [Some discussion about mathematics assignment assessment in primary education]	小学生(教学实践) [Primary school students (teaching practice)]
236	郑凌川 [Zheng, L.]	2012	浅谈新课程理念下小学数学课堂教学评价 [Some discussion mathematics classroom teaching and learning assessment in primary education under the guidance of new curriculum reform]	科学咨询(教育科研) [Scientific consultation (educational research)]
237	郑华恒 [Zheng, H.]	2012	让即时评价充满生命的张力 – 小学数学课堂即时评价的思考与实践 [Make immediate assessment full of vitality: practice of and reflection on mathematics classroom immediate assessment in primary education]	教育教学论坛 [Forum of education and teaching]
238	郭小玉 [Guo, X.]	2012	延时评价在数学教学中的作用 [The function of delayed assessment in mathematics teaching]	广西教育 [Education in Guangxi]
239	郭志禹 [Guo, Z.]	2012	拓宽反馈途径优化教学信息 [Broaden the approaches to provide feedback and optimize teaching information]	小学教学参考 [Reference of primary school teaching]
240	郭忠勤 [Guo, Z.]	2012	如何恰如其分地评价学生的数学作业 [How to assess students' mathematics assignment appropriately]	教育革新 [Educational innovation]
241	钱丽华 [Qian, L.]	2012	恰当运用教学评价,促进学生的可持续发展 [Using teaching assessment appropriately to promote students' sustainable development]	考试周刊 [Weekly examination]
242	闫广富 [Yan, G.]	2012	数学课堂教学中的评价手段 [Assessment methods in mathematics classroom teaching]	黑河教育 [Education in Heihe]
243	闻安凤 [Wen, A.]	2012	扬起评价之帆助低年段学生快乐起航 – 浅谈低年段数学课堂评价的三大策略 [Set the sail of assessment to help primary students in lower grades to start: three strategies of using classroom assessment in mathematics]	文理导航(下旬) [Guidance of arts and science]
244	陈光美 [Chen, G.]	2012	对小学数学课堂教学评价的质性研究 [Qualitative research on mathematics classroom teaching and learning assessment in primary education]	数学大世界(教学导向) [The world of mathematics (teaching guide)]

#	Author	Year	Title	Journal
245	陈凤霞 [Chen, F.]	2012	良言一句暖三冬 – 浅谈课堂评价语言的作用 [One golden saying can warm for three winters: the function of classroom assessment language]	现代阅读(教育版) [Modern reading (education)]
246	陈小丽 [Chen, X.]	2012	关注差异，进行多元评价 [Focus on differences and conduct the multi-actor system of assessment]	教育教学论 [Education and didactics]
247	陈小霜 [Chen, X.]	2012	小学数学课堂评价应做到"三不" ["Three things not to do" in classroom assessment in primary school mathematics education]	广西教育 [Education in Guangxi]
248	陈峥嵘 [Chen, Z.]	2012	对小学数学课堂评价用语的研究 [Research on mathematics classroom assessment language in primary education]	课程教育研究 [Educational research on curriculum]
249	陈微云 [Chen, W.]	2012	学习性评价在小学数学课堂中的实践 [Practice of learning assessment in primary school mathematics classroom]	考试周刊 [Weekly examination]
250	陈栋 [Chen, D.]	2012	坚持生本化评价理念实施多元化评价方式 – 浅谈小学数学课堂教学评价的人本化问题 [Follow the assessment idea of student-centered and conduct a multi-actor system of assessment: discussion about the problem of student-centered in classroom teaching and learning assessment in mathematics in primary schools.]	小学时代(教师) [Primary times (teachers)]
251	陈秀娟 [Chen, X.]	2012	让评价用语在数学课堂上飞扬 [Make assessment language fly in mathematics classes]	吉林教育 [Education in Jilin]
252	陈芳 [Chen, F.]	2012	多元主体评价促进学生发展 [Establish a multi-actor system of assessment to promote students' development]	新课程研究(上旬刊) [Research in new curriculum]
253	陈计宏 [Chen, J.]	2012	小学数学课堂评价如何突破 [How to make a breakthrough in classroom assessment in primary school mathematics education]	考试周刊 [Weekly examination]
254	韩秀娟 [Han, X.]	2012	小学数学教学中即时评价的现象及对策 [Phenomenon and strategies of immediate assessment in mathematics teaching in primary education]	小学教学参考 [Reference of primary school teaching]
255	项薇薇 [Xiang, W.]	2012	莫让时机顺水流 – 浅谈数学课堂即时评价 [Do not make the opportunity missing: some discussion about immediate assessment in mathematics classroom]	中小学电教(下) [Primary and middle school educational technology]
256	顾明 [Gu, M.]	2012	数学教学应在有效理答上着力 [Mathematics teaching should pay more attention to effective questioning and answering]	成才之路 [The road of being a useful person]

#	Author	Year	Title	Journal
257	顾维敏 [Gu, W.]	2012	评价 – 学生发展的羽翼 – 小学数学作业评价策略的实践研究 [Assessment as the wings of students' development: practice of and research on using assignment assessment in primary school mathematics education]	数学学习与研究 [Mathematics learning and research]
258	马秀玲 [Ma, X.]	2012	多媒体在数学评价中的运用 [Use multimedia in assessment in mathematics]	中小学电教(下) [Primary and middle school educational technology]
259	马雪飞 [Ma, X.]	2012	谈小学数学课堂教学即时评价的误区与对策 [Discussion about misunderstanding and strategies of using immediate assessment in primary school mathematics teaching]	新课程学习(上) [Learning of new curriculum]
260	高企丰 [Gao, Q.]	2012	促进有效课堂交往,提升小学数学课堂教学实效 [Promote effective classroom exchange to facilitate effectiveness of classroom teaching in primary school mathematics education]	新课程(下) [New Curriculum]
261	黄传侠 [Huang, C.]	2012	小学数学教师如何在课堂上实施有效的评价语言 [How primary school mathematics teachers should use effective assessment language in classroom]	新课程(上) [New Curriculum]
262	黄日红 [Huang, R.]	2012	如何提高数学课堂教学的有效性 [How to improve effectiveness of mathematics teaching in classroom]	新课程(小学) [New curriculum (primary education)]
263	黄晨莹 [Huang, C.]	2012	"激励"不是课堂评价的全部 – 小学数学课堂评价策略初探 ["Stimulation" is not all of classroom assessment: exploration of the strategies of using classroom assessment in primary school mathematics education]	新课程(上) [New Curriculum]
264	黄荣元 [Huang, R.]	2012	浅析如何改革小学数学作业的评价方式 [Analysis of how to reform assessment methods in assessing mathematics assignment in primary education]	小学教学参考 [Reference of primary school teaching]
265	黄莉 [Huang, L.]	2012	如何让大班化数学课堂活起来 – 从差异与协同的角度思考课堂教学 [How to make mathematics classroom with a number of students lively: considering classroom teaching from the perspectives of having both difference and things in common]	课程教育研究 [Educational research on curriculum]
266	龙强 [Long, Q.]	2012	新课改中数学教学行为的转变 [The change in mathematics teaching]	科教文汇(下旬刊) [The science education article collects]

Appendix B

Coding framework for teacher-authors' conceptions of classroom assessment added with examples (Chapter 2)

Category	Subcategory	Example
Purpose	Checking students' understanding	#63: The teacher assesses where her students get stuck or have misunderstandings by asking them to discuss the characteristics of a fraction which can be written as a decimal.
	Stimulating students to learn	#5: The teacher states that the purpose of assessment is to facilitate students' development. #8: The teacher states that it is necessary for teachers to use assessment to train and develop students' concentration and other good learning habits.
	Informing teachers' instructional decision-making	#51: The teacher states that one of the fundamental purposes of assessment is to facilitate teachers' reflection and improvement of instruction. #7: The teacher gives the quotation about the purpose of assessment in mathematics curriculum standards, including improving teachers' instruction.
	Establishing a harmonious classroom environment	#8: The teacher states that classroom assessment is a key factor to establish a good learning environment.
	Promoting students' confidence	#93: The teacher organizes the peer-assessment in order to establish a student's confidence.
Content	Basic knowledge and skills	#11: The teacher assesses her students' mental calculation by giving exercises.
	Mathematical thinking and problem solving	#76: The teacher assesses whether her students can recognize the similarity or differences between six groups of additions problems.

Category	Subcategory	Example
Content	Mathematical and learning attitude	#23: The teacher states that teachers should observe whether students are good at cooperating with others, whether the students are participating in learning actively.
Assessor	Teachers	#2: The teacher assesses her students by asking questions.
	Students (including assessing themselves and peers)	#5: The teacher describes that through using self- and peer-assessment, students actively take part in assessment and the function of classroom assessment is magnified. #6: The teachers describe how they let their students correct their homework in groups.
	Parents	#16: The teacher describes that parents are asked to fill in a form in which they have to describe how their children do homework. #33: The teacher describes that parents assess their children's homework once a week and give written feedback as well.
Method	Observation	#56: The teacher describes that she walks around and observes how her students are doing the assignments.
	Questioning	#27: The teacher shows the algorithm of a problem asks her students: "Is 23+143=373 correct?"
	Classroom discussion	#64: The teacher asks the students to discuss in class the different strategies to solve the problem 510-380.
	In-class assignment	#11: The teacher gives her students ten mental calculation problems in class.
	After-class assignment	#3: After a lesson about money, the teacher asks the students to go shopping and report the price of the goods.

Category	Subcategory		Example
Method	Presentation		#51: The teacher asks a student to show how he uses the abacus to calculate addition problems.
	Portfolio		#23: The teacher describes what materials can be put into the portfolios.
	Quiz/written test		#60: The teacher suggests using written tests to check students' learning.
	Mathematics diary		#43: The teacher describes how she guides her students to write a diary by reporting their understanding or finding related to mathematics.
Feedback	Focus	Task-related	#36: The teacher states that feedback should make clear to students what is right or wrong in their answers and why. #43: The teacher uses feedback to point out her students' mistakes in their homework by writing "Is the order of operations correct?"
		Process-related	#41: The teacher gives examples of feedback related to solving problems in general, such as "I believe if you think for a while, you can find the correct answer", "Please recalculate the problem more carefully" and "Won't it be better if you can write the homework in a good way?". #43: The teacher explains that written feedback should be used to help students develop a good habit of doing homework.
		Person-related	#10: The teacher gives feedback such as "You are great!" and "[…] You are a great little mathematician!" #29: The teacher describes that he gives his students simple but warm written feedback, such as "You are very smart!"

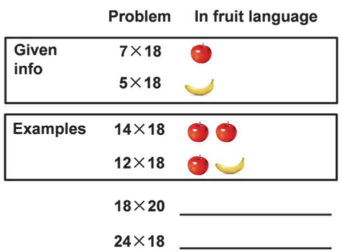
Category	Subcategory		Example
Feedback	Nature	Only positive	#8: The teacher states that teachers should not be parsimonious when praising students. He gives feedbacks like "You are the one who listens most carefully" and "You are really good at listening!" #11: The teacher praises her students when they perform well in order to build up their confidence. Examples of the feedback given by her are: "You are extraordinary!", "I believe you can do it!" and body language like smiling.
		Balanced (positive and what needs to be improved)	#5: The teacher agrees to provide instant praise to students when they make progress. Furthermore, he makes it clear that besides encouraging and respecting students, it is necessary to correct them to make them learn from frustrations and know where and how to improve. #43: The teacher describes how she praises her students to arouse their learning potential. Also, she points out how her students can improve by giving written feedback: "The order of operations! See, you made a mistake because of carelessness"; and "Think over what should be considered first."
	Mode	Verbal	#2: The teacher gives feedback by saying "Your method is very good!" and "You are very brilliant!" #4: The teacher explains that he/she gives the students feedback on their homework face to face.
		Written	#4: The teacher gives written feedback by using mark, star and face together. For example: "A '4' means neat work, the maximum value is '5'; '☆☆☆' means moderate accuracy, five times a ☆means no mistake; ☺ means a good attitude, 😐 and ☹ are other options for assessing students' attitude." #13: The teacher gives written comments to his students, such as "Is the order of operations correct?" and "Please recalculate the result more carefully."

Category	Subcategory		Example
Feedback	Mode	Body language	#5: The teacher describes why body language is important #11: The teacher combines body language with verbal feedback. She smiles to a student who answers in a low voice while saying “If you speak louder, all the students in our class can hear you clearly. Isn’t that better?”
		Material incentives	#21: The teacher describes how to use an accumulative reward system of incentives, which is called “Little tree as a companion of growing up.” She gives each student an apple tree that only consists of a trunk and empty branches. The students can earn stars; stars can be exchanged for apples which they can hang in their own tree.
	Timing	Immediate (during class)	#2: The teacher gives verbal feedback in class. #78: The teacher corrects her students’ work directly after they finish the exercises in class.
		Delayed (after class)	#4: The teacher explains that he/she gives the students face-to-face feedback on their homework after class. #22: The teacher explains how he/she gives written feedback on the students’ homework after class.
Teaching adaptation			#38: The teacher gives supplementary exercises immediately after finding her students could not answer the questions correctly. #86: The teacher describes to be required to analyze students’ results of written tests and students’ mistakes in order to adapt further teaching.

Appendix C

The eight CATs for multiplication of two-digit numbers (Chapter 5)

<table>
<tr><th>Title</th><th>Format</th><th>Example</th><th>Purpose</th></tr>
<tr><td>Family problems (CAT-1)</td><td>Red/Green cards</td><td>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</td><td>Assessing whether students can recognize analogous problems and are aware of the relationship among the results of these problems</td></tr>
<tr><td>How many approximately? (CAT-2)</td><td>Red/Green cards</td><td>Here you see five bags with a different weight.
How many kilograms approximately in one bag?
Do you think the answer is 30, 40 or 35?
(Yes- Green card ; No- Red card)
<table><tr><th>Bag 1</th><th>Bag 2</th><th>Bag 3</th><th>Bag 4</th><th>Bag 5</th></tr><tr><td>33 kg</td><td>34 kg</td><td>37 kg</td><td>37 kg</td><td>36 kg</td></tr></table></td><td>Assessing whether students can find the approximate average of a series of numbers</td></tr>
<tr><td>Breaking down a multiplication (CAT-3)</td><td>Worksheet</td><td>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) 3×4 and 3×20 and __×4 and __×20
b) 4×__ and 4×__ and __×3 and 20×50
c) 20×50 and 20×__ and 4×50 and __×3</td><td>Assessing whether students understand how the multiplication algorithm works</td></tr>
<tr><td>Completing the ratio table (CAT-4)</td><td>Red/Green cards</td><td>You are NOT allowed to calculate how many pencils there are in one box. Can you still find out how many pencils there are in the following boxes?
(Yes- Green card ; No- Red card)
<table><tr><th>Number of boxes</th><th>6</th><th>11</th><th>12</th><th>13</th><th>17</th><th>18</th><th>22</th><th>23</th></tr><tr><td>Number of pencils</td><td>72</td><td>132</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></td><td>Assessing whether students can make use of the ratio table by reasoning horizontally</td></tr>
</table>

Title	Format	Example	Purpose
Bigger or smaller? (CAT-5)	Red/Green cards	Are the statements correct? (Yes- Green card ; No- Red card) a) The product of **24×45** is smaller than **1200**. b) The product of **18×32** is bigger than **600**.	Assessing whether students can estimate the product by reasoning
Quick check of answers (CAT-6)	Red/Green cards	Here you see multiplication problems with their answers. Do you see a mistake? (Yes- Green card ; No- Red card) 50×200=1000 37×34=1250 23×38=874	Assessing whether students can quickly check the correctness of the result of multiplication problems without performing the algorithm
Fruit language (CAT-7)	Worksheet	On this worksheet you have to rewrite the problems in fruit language. Use the given info and look at the examples.  	Assessing whether students can use the associative and distributive property of multiplication to restructure a multiplication problem
Solving problems without using the algorithm (CAT-8)	Worksheet	Here you see a multiplication problem. You have to solve it WITHOUT using the algorithm. Write down how you solved it. [ ] In this way I solved the problem / Answer 59×62	Assessing whether students have a deep understanding of the multiplication operation and whether they have, instead of the algorithm, other strategies available to solve multiplication problems

Samenvatting

中文概要

Acknowledgements

Curriculum Vitae

List of publications related to this thesis

List of presentations related to this thesis

FI Scientific Library

ICO Dissertation Series

Samenvatting

In China, waar al sinds jaar en dag een cultuur heerst van het afnemen van examens en toetsen gericht op selectie, is in 2001 een nationaal onderwijshervormingsplan ingevoerd, met als doel meer aandacht te geven aan het gebruik van toetsen om het leren en het onderwijzen te ondersteunen. Een decennium later zijn er echter nog maar weinig studies uitgevoerd naar de praktijk van het toetsen in de klassen van leerkrachten rekenen-wiskunde in het basisonderwijs. Met deze achtergrond, is in 2012, het *Improving Classroom Assessment in China* (ICA-C) project opgezet, bestaande uit vier deelstudies die in de respectievelijke hoofdstukken van dit proefschrift zijn beschreven. Het Chinese ICA-project was een vervolg op het Nederlandse ICA-project.

Een belangrijk doel van het ICA-C-project was de huidige situatie van de toetspraktijk en ideeën daarover van Chinese reken-wiskundeleerkrachten in het basisonderwijs in kaart te brengen. Hierbij richtten we ons met name op het toetsen in de klas door de leerkracht. Dit wordt *classroom assessment* genoemd en omvat het (formatief) toetsen door de leerkracht voor, tijdens en na het lesgeven, waarbij de leerkracht gebruik maakt van allerlei activiteiten die meer zicht geven op het kennen en kunnen van de leerlingen. Door de informatie die op deze manier verkregen is, kan de leerkracht vervolgens zodanig het verdere onderwijs aanpassen dat het aansluit op de leerbehoeften van de leerlingen.

Om te onderzoeken hoe Chinese leerkrachten over *classroom assessment* denken en wat ze in dit verband doen is een review uitgevoerd van artikelen in nationale vaktijdschriften die door leerkrachten over hun toetsactiviteiten en –ideeën zijn geschreven. Van deze review wordt verslag gedaan in Hoofdstuk 2. De artikelen die zijn geanalyseerd zijn ontleend aan de *China National Knowledge Infrastructure*-database. Bij het zoekproces hebben we ons beperkt tot de publicatiejaren 2011 en 2012. Dit leidde tot 266 door leerkrachten geschreven artikelen over classroom assessment in het reken-wiskundeonderwijs op de basisschool. Om deze artikelen te analyseren is een codeerschema ontwikkeld waarbij we zijn uitgegaan van de toetsrichtlijnen die in het Chinese rekenen-wiskundecurriculum zijn opgenomen. Zo maakten kenmerken als het doel van de toets, de inhoud van de toets, degene die toetst, de toetsmethode, het geven van feedback en het gebruik van toetsresultaten voor het aanpassen van de verdere instructie deel uit van het codeerschema. Bij het besluiten over verdere subcategorieën werden de thema's die de auteurs van de artikelen aanstipten meegenomen. Uit de resultaten bleek dat de auteurs met name reflecteerden op: het doel van de toets (in 49% van de 266 artikelen), de te toetsen leerinhouden (70%), degene die toetst (100%) en de gebruikte toetsmethoden (78%). In het bijzonder werd aan het geven van feedback veel aandacht geschonken: in 198 artikelen (74%) ging minstens een kwart van de tekst over feedback, waarbij in 64 van deze artikelen classroom assessment en feedback als synoniemen werden gezien door de auteurs. Opvallend was dat de leerkrachten in slechts negen artikelen (3%) een aanpassing van de instructie op basis van de verkregen informatie beschreven. In 142 artikelen (53%) beschreven de leerkrachten dat ze toetsten volgens de officiële toetsrichtlijnen, door dan wel te citeren uit deze richtlijnen of ze te parafraseren. Zo bleken ook de ideeën van leerkrachten over toetsen, zoals beschreven in hun artikelen, en de toetsrichtlijnen veel overeenkomsten te vertonen, behalve dan het aanpassen van de instructie op basis van de verkregen informatie.

In een tweede studie hebben we met een vragenlijst bij Chinese basisschoolleerkrachten aanvullende gegevens verzameld over hun toetspraktijk bij rekenen-wiskunde en hun ideeën daarover. Van dit vragenlijstonderzoek wordt verslag gedaan in Hoofdstuk 3. In deze vragenlijst werden leerkrachten gevraagd om hun toetspraktijk bij rekenen-wiskunde en hun ideeën daarover te beschrijven, en deze was gebaseerd op een eerder gebruikte vragenlijst in Nederland en aangepast aan de Chinese context. In totaal werden er 30 vragen gesteld over de achtergrond van leerkrachten, hun algemene onderwijspraktijk en hun toetspraktijk en ideeën daarover. Van 1101 leerkrachten uit 12 Chinese provincies en regio's werden ingevulde vragenlijsten ontvangen en geanalyseerd. Door gebruik te maken van exploratieve factoranalyses werd de onderliggende structuur van de vragenlijst blootgelegd, en die bestond uit acht factoren: *(1) toetsen om algemene beslissingen over instructie te nemen, (2) toetsen om specifieke beslissingen over instructie te nemen, (3) toetsmethoden, (4) diversiteit van probleemtypes, (5) belang van het toetsen van kennis en vaardigheden, (6) belang van het toetsen van bijzondere vaardigheden, (7) gepercipieerde nut en (8) acceptatie van toetsing.* Met een latente klasse analyse werden drie toetsprofielen van leerkrachten bepaald. Door naar deze latente klassen van leerkrachten te kijken met de acht factoren van de vragenlijst konden de drie toetsprofielen van leerkrachten geformuleerd worden.

De meeste leerkrachten hoorden bij het *Mainstream assessors* profiel (53% van de leerkrachten). Deze leerkrachten waren gemiddeld in het gebruik van toetsen, ze scoorden namelijk rondom de gemiddelde score op de meeste factoren en relatief hoog op *acceptatie van toetsing*. Om precies te zijn rapporteerden deze leerkrachten dat ze de verschillende toetsmethoden voor verschillende doelen met een gemiddelde frequentie gebruikten. Om leerlingen te toetsen gebruikten ze verschillende probleemtypes. Ook onderstreepten ze in het algemeen het belang van het toetsen van verschillende soorten kennis en vaardigheden en vonden toetsen nuttig ter ondersteuning van het leren en onderwijzen. Daarbovenop waren deze leerkrachten, van de leerkrachten met de drie toetsprofielen degenen die toetsen het meest accepteerden in hun praktijk. De tweede groep leerkrachten hoorden bij het *Enthusiastic assessors* profiel (22%), zij hadden bovengemiddelde scores op de verschillende factoren. Zij rapporteerden heel vaak verschillende toetsmethoden voor verschillende doeleinden te gebruiken, en onderstreepten het belang van het toetsen van kennis en vaardigheden, ook vonden ze toetsen zeer nuttig. De overige leerkrachten (25%) waren de *Unenthusiastic assessors*. Deze leerkrachten scoorden op bijna alle factoren ruim onder gemiddeld, wat betekent dat ze toetsen niet vaak en niet voor verschillende doeleinden gebruiken, en ook niet belangrijk of nuttig vinden. Na deze indeling van leerkrachten in toetsprofielen te hebben verkregen, en zo meer zicht te hebben op de huidige situatie van *classroom assessment* bij rekenen-wiskunde op de basisschool, hebben we ons op het uitzoeken van hoe deze praktijk eventueel verbeterd zou kunnen worden gericht.

Het andere doel van het ICA-C-project was om uit te zoeken in hoeverre de formatieve toetspraktijk van Chinese reken-wiskundeleerkrachten in het basisonderwijs verbeterd kan worden, waarbij we ons in het bijzonder richtten op het gebruik van *classroom assessment techniques* (CATs). CATs zijn korte toetsactiviteiten die dicht tegen het curriculum zitten en die de leerkracht initieert. Leerkrachten gebruiken deze CATs in hun dagelijkse lespraktijk om het begrip van bepaalde reken-wiskundige concepten of vaardigheden die leerlingen hebben te weten te komen. In de derde studie is het gebruik van zulke CATs door Chinese reken-wiskundeleerkrachten in het basisonderwijs verkend. Van deze case study wordt verslag gedaan in Hoofdstuk 4. Zes vrouwelijke reken-wiskundeleerkrachten van groep 5 en hun 216 leerlingen van twee verschillende basisscholen in Nanjing, China, deden mee aan deze studie. Op basis van de gebruikte lesmethode hebben we een serie CATs ontwikkeld om het begrip van leerlingen van delingen van driecijferige getallen door getallen van een cijfer te toetsen. Om de leerkrachten te ondersteunen in het gebruik van de CATs hebben we hier ook een handleiding bij geschreven waarin het doel van iedere CAT en suggesties voor de uitvoering beschreven worden. Daarnaast kregen de leerkrachten te horen dat het ze vrijstond om de CATs aan te passen aan en in hun eigen onderwijspraktijk. Gedurende de twee weken waarin dit deelonderwerp werd onderwezen, werden vier bijeenkomsten van een uur georganiseerd die dienden om de leerkrachten hierin te ondersteunen. In deze bijeenkomsten werd besproken hoe de CATs gebruikt konden worden en hoe de leerkrachten ze gebruikt hadden in hun eigen klassen. Gegevens werden verzameld over het gebruik van de CATs door leerkrachten te interviewen en feedbackformulieren en eindrapporten te verzamelen. Daarbij werden ook lessen geobserveerd en leerlingenwerk bekeken, om zo de informatie die de leerkrachten zelf aanleverden aan te vullen.

Uit de analyse van deze gegevens bleek dat de leerkrachten de CATs in hun lespraktijk gebruikten ondanks dat ze nieuw voor hen waren. De leerkrachten pasten de CATs aan zodat ze bij hun vooropgestelde lesplanning pasten. De aanpassingen waren meestal organisatorisch van aard, zoals bijvoorbeeld het aantal taken of de timing van een CAT aanpassen, of het stap-voor-stap uitleggen van de procedure die in de CAT gevraagd zou worden. Door gebruik te maken van de CATs kregen leerkrachten nieuwe informatie over het begrip van hun leerlingen. Leerkrachten stelden de CATs het meest op prijs waarin vragen werden voorgelegd aan de leerlingen over lesstof die niet volledig door het lesboek was uitgelegd, waardoor de leerkrachten echt nieuwe informatie uit de CATs konden halen. Daarnaast waren leerkrachten met name gecharmeerd van de werkvorm met de rode en groene kaartjes, waarmee ze snel informatie over hun leerlingen kunnen vergaren. Desalniettemin bleken leerkrachten de informatie die ze dankzij de CATs hadden vergaard niet te gebruiken om hun verdere onderwijs en instructie aan te passen in de daaropvolgende lessen. Wel bleken ze de informatie enigszins te gebruiken, namelijk direct na of tijdens het uitvoeren van een CAT gaven leerkrachten vaak extra uitleg aan leerlingen zodat ze tot het goede antwoord zouden komen. Daarnaast gebruikten leerkrachten de handleiding bij de CATs om hun onderwijs van tevoren al aan te passen. Zo werden de CATs gebruikt als extra oefeningen in de geplande lessen van de leerkrachten in plaats van als toetsmoment. In het algemeen vonden leerkrachten dat de CATs goed te gebruiken waren, informatie over het leren van de leerlingen gaf, efficiënt daarin waren en ook nuttig konden zijn om leerlingen bij het toetsen te betrekken.

In de vierde en laatste studie (zie Hoofdstuk 5 van dit proefschrift) is onderzocht welke nieuwe inzichten leerkrachten krijgen in het begrip en de vaardigheden van leerlingen door het gebruik van de CATs. De opzet van deze studie vertoonde veel gelijkenis met de derde studie (zie Hoofdstuk 4), met dien verstande dat een grotere groep leerkrachten meedeed en een ander onderwerp was gekozen. Om precies te zijn kregen 25 leerkrachten van rekenen-wiskunde in groep 5 van negen scholen in Nanjing acht CATs waarmee ze het begrip van vermenigvuldigen met tweecijferige getallen van hun leerlingen konden toetsen. Hiertoe was weer een handleiding opgesteld voor bij het gebruik van de CATs en werden twee bijeenkomsten van twee uur georganiseerd waarin de leerkrachten werden ondersteund bij de implementatie van de CATs. Leerkrachten vulden feedbackformulieren in na het uitvoeren van de CATs en schreven een eindrapportage over hun gebruik van de CATs. Deze beschrijvingen zijn gebruikt om te determineren in hoeverre ze nieuwe inzichten ontwikkelden over het reken-wiskundig begrip van hun leerlingen door het gebruik van de CATs. Een beschrijving werd gecodeerd als “getuigt van nieuw inzicht” als de leerkracht de reken-wiskundige inhoud die in een CAT getoetst werd beschreef, en ofwel specifieke informatie over de leerlingen beschreef ofwel de nieuwigheid van de verkregen informatie beschreef ofwel een passende aanpassing van de verdere instructie beschreef. De scores van hun leerlingen op drie districttoetsen rekenen-wiskunde werden geanalyseerd om te verkennen of het verkrijgen van nieuwe inzichten door de leerkrachten samenhing met veranderingen in de prestaties van hun leerlingen op het gebied van rekenen-wiskunde.

In totaal werden 193 beschrijvingen van leerkrachten over hun gebruik van de CATs verzameld, waarvan 57 beschrijvingen duidelijk getuigden van nieuwe ontwikkelde inzichten bij de leerkracht. Afhankelijk van het aantal CATs waarmee leerkrachten de inzichten hadden verkregen werden ze in drie groepen gedeeld voor de verdere analyse. Vijf leerkrachten behoorden tot de *Hoog Inzicht*-groep, aangezien zij nieuwe inzichten bleken te hebben verworven dankzij vijf of meer CATs, veertien leerkrachten hoorden bij de *Enig Inzicht*-groep, zij hadden bij drie of minder CATs nieuwe inzichten verworven en de overige zes leerkrachten hoorden bij de *Geen Inzicht*-groep aangezien uit hun beschrijvingen niet bleek dat ze nieuwe inzichten hadden verworven. De leerkrachten in de *Hoog Inzicht*-groep gaven aan het idee van de CATs in te zien en gebruikten ze echt om meer te weten te komen over het leren van hun leerlingen. In het algemeen vonden deze leerkrachten de CATs waarin de taken duidelijk anders waren dan in het lesboek het meest onthullend. Daarentegen rapporteerden de leerkrachten in de *Geen Inzicht*-groep vaak simpelweg in algemene termen of hun leerlingen de taken op konden lossen of niet. Verschillende van deze leerkrachten vonden dat de CATs te veel op het lesboek leken en wilden niet nog eens herhalen wat ze al eerder hadden onderwezen, anderen onderstreepten echter dat de CATs juist te veel verschilden van het lesboek. Als naar de testscores van de leerlingen op de drie toetsen werd gekeken bleek op de eerste plaats dat de leerlingen op alle drie de districttoetsen heel hoog scoorden. Meer dan twee derde van de leerlingen had op deze toetsen een goedscore van 85% of hoger. Na de interventie was er een (heel) kleine toename in de scores van de leerlingen van leerkrachten in de *Hoog Inzicht*-groep en een (heel) kleine afname bij leerlingen van leerkrachten in de *Geen Inzicht*-groep. Met een covariantieanalyse waarin de gestandaardiseerde voortoets-score als covariaat was gebruikt en inzicht groep als factor, bleken deze effecten van verworven inzichten van leerkrachten op de reken-wiskundeprestaties van de leerlingen niet significant te zijn.

In Hoofdstuk 6 wordt een overzicht van de bevindingen uit de vier studies van het ICA-C-project beschreven evenals suggesties gegeven voor de toetspraktijk en verder onderzoek naar classroom assessment. De resultaten van het ICA-C-project duiden erop dat de deelnemende Chinese leerkrachten rekenen-wiskunde in het algemeen achter het idee stonden dat toetsen gebruikt worden om het onderwijsleerproces te verbeteren. Overeenkomstig de onderwijsrichtlijnen van China, gebruikten ze verschillende toetsmethoden om verschillende typen kennis en vaardigheden te toetsen voor een verscheidenheid aan doeleinden. Echter bleken leerkrachten zelden de zo verkregen informatie te gebruiken voor het aanpassen van hun verdere instructie en onderwijs. CATs bleken goed uitvoerbaar en potentieel nuttig voor het verbeteren van het *classroom assessment* van leerkrachten. Door te toetsen vanuit een andere invalshoek dan wat door het lesboek wordt voorbereid kunnen CATs leerkrachten toegang geven tot nieuwe inzichten in het reken-wiskundig begrip van hun leerlingen. Desalniettemin is ook gebleken dat het gebruik van CATs voor Chinese leerkrachten nog wel een uitdaging is. Meer ondersteuning en begeleiding van beleidsmakers, onderwijsonderzoekers en lerarenopleiders is nodig voor leerkrachten om hun *classroom assessment* optimaal te benutten.

中文概要

中国有着根深蒂固的考试文化。自2001年起，评价改革在中国内地大力推行，倡导发挥评价改进教师教学实践和促进学生学习的功能。然而，十多年过后，只有为数不多的研究聚焦于小学数学教师的课堂评价行为。在此背景下，研究项目"**促进在中国的课堂评价**"(Improving Classroom Assessment in China，简称ICA-C)于2012年正式开启。该项目是荷兰"促进课堂评价"(Improving Classroom Assessment，简称ICA)研究项目的延续和扩展。

ICA-C项目有两个研究目标。第一个研究目标是对中国小学数学教师对课堂评价理解和实践的现状进行勾勒。**课堂评价**，即由教师主导实施的形成性评价，涵盖教师为使教学决策充分满足学生当下的学习需求而在日常教学中持续开展的多种活动。为了调查中国小学数学教师如何思考、运用课堂评价，我们设计了两个研究子课题。在第一个子课题中，我们针对教师写作发表的、着重讨论课堂评价的期刊文章进行综述；在第二个子课题中，我们借助调查问卷进一步收集数据调研教师对课堂评价的理解和实践。

ICA-C项目的第二个研究目标是探究运用课堂评价技术促进中国小学数学教师课堂评价的可能性。**课堂评价技术** (Classroom assessment techniques，简称CATs) 是由教师发起的、与教科书密切相关且能够在日常教学中使用的简短评价活动，旨在揭示学生对特定数学概念或技能的掌握情况。为了调查中国小学数学教师如何使用CATs，能够借助CATs洞察到哪些有关学生数学理解的信息，我们设计了另外两个子课题。简言之，ICA-C研究项目共包括四个子课题，分别在博士论文的第二章到第五章详细阐述。

在第二章中，我们针对小学数学教师对课堂评价的认识理解进行综述研究。综述的对象是小学数学教师写作发表的关于课堂评价的期刊文章，其内容通常包括教师对个人课堂评价实践的记录描述和对习得评价经验的讨论总结。文章检索在中国知网期刊数据库中进行，筛选后266篇发表于2011年和2012年的文章符合研究要求。为了分析这些文章，我们制定了编

码框架。编码框架中的主要类别与《中国义务教育数学课程标准(2011版)》中提到的评价的几个主要方面一致，包括评价目的、评价内容、评价主体、评价方式、提供反馈和利用评价结果改进教学。在进一步设定每个主要类别的子类别时，也参考了教师在文章中讨论的内容。

研究结果显示，这些教师对评价目的(占全部266篇文章的49%)、评价内容(70%)、评价主体(100%)和评价方式(78%)进行了反思。需要特别指出的是，提供反馈是教师反思讨论的重点。在198篇文章(74%)中，教师用至少四分之一的篇幅讨论如何提供反馈；其中64篇文章的作者认为课堂评价等同于反馈。与上述结果形成鲜明对比的是，仅有9篇文章(3%)的作者介绍或解释了其如何根据获取的评价信息来改进教学的。将教师对课堂评价的理解(经由分析其发表的文章得到)和数学课程标准中提倡的相关评价内容比较，除了利用评价结果改进教学这一方面，其余五个方面都能够很好地对应。另外，在142篇文章(53%)中，教师或者明确指出自己的评价活动是依据课标中的评价建议进行的，或者引用、转述评价建议的原文。

在第三章中，我们开展了大规模问卷调查研究，旨在进一步勾勒中国小学数学教师对评价理解。我们选用在荷兰ICA项目中设计使用的调查教师评价实践和评价信念的问卷，并根据中国大陆的实际情况对问卷进行调整。调整后的问卷包括30个问题，收集教师的背景情况、教学实践、评价实践和评价信念等信息。最终，研究分析了来自中国大陆12个省市自治区的1101位小学数学教师的问卷。通过探索性因子分析法，我们揭示了问卷的内在本质结构并确定了八个因子：(1)用于一般性教学决策的评价目的，(2)用于具体教学决策的评价目的，(3)评价方式，(4)评价问题形式的多样性，(5)评价知识技能的重要性，(6)评价非课标重点强调技能的重要性，(7)评价的效用，(8)对评价的接受程度。通过潜在类别分析，我们确定了三类评价人群。通过分析这三类评价人群在上述8个因子上的不同表现，揭示了教师对评价的不同看法。

大部分教师属于**主流评价者**(53.1%)，对评价的使用情况适中。这类教师在大多数因子上的得分都接近平均值；但在“对评价的接受程度”这一因子上得分较高。具体地说，属于主流评价者的教师对评价的使用频率处于

中等水平，会使用几种评价方式来达到不同的评价目的来帮助其日常教学决策。评价学生时，他们会使用不同形式的问题。另外，主流评价者基本认同对学生各种知识技能进行评价的重要性，认可评价对教师教学和学生学习的促进作用。与另外两类教师人群相比，主流评价者在实践中对评价的接受程度也最高。部分教师属于**热情评价者**(21.7%)，他们在绝大多数因子上的得分都高于平均值。这类教师会频繁地使用各种不同的评价方式来实现多种评价目的。热情评价者高度赞同对学生的各种知识技能进行评价的重要性，认为评价非常有用。其余的教师(25.2%)属于**冷淡评价者**。这类教师在几乎所有因子上的得分都远低于平均值，说明他们对评价的使用缺乏明确的目标，不规律，也不认为评价是重要或有用的。在对中国大陆小学数学课堂评价的现状有了一定了解后，我们的研究重点转向如何提高教师的课堂评价实践。

在第四章中，我们介绍了针对小学数学教师使用**课堂评价技术**的探索性研究。六位来自中国南京两所小学的三年级女教师及其216名学生参与了此项研究。根据教师们使用的教科书(苏教版)，我们设计了一系列课堂评价技术用于评价学生对三位数除以一位数的除法运算的理解。为了帮助教师更好地理解和使用课堂评价技术，我们编写了《教师CATs使用手册》详细介绍每个技术的使用目的和操作建议。同时，我们也向教师说明，他们无需按部就班地完成手册中的每个步骤，可以根据自己班级的教学实情调整技术的使用。在两周的除法教学中，我们组织了四次时长为1小时的会议，和教师讨论如何使用技术或者由教师分享其使用技术的体会。关于教师使用课堂评价技术的信息，我们主要通过教师访谈、教师反馈表和教师使用小结的方式进行收集。必要情况下，我们也会使用课堂观察记录和学生工作单来补充或佐证教师汇报的情况。

根据教师的反馈，我们发现：虽然各位教师是初次使用课堂评价技术，但是他们能够很容易地将课堂评价技术应用于教学实践。通常教师会对技术进行调整，使技术与其原有的教学计划相容。这些调整包括减少课堂评价技术中题目的数量，改变技术使用的时间，在实施评价前或者进行评价中直接教授学生如何解决题目。通过使用技术，教师对学生的数学理解有了新的认识。具体地说，教师很欣赏课堂评价技术的提问方式，它们与教科

书中问题的呈现方式不同，对学生提出了挑战，从而极具揭示意义。另外，大部分教师都喜欢红绿卡片形式的课堂评价技术，因为它们能够快速反映学生的理解情况。教师会在阅读《教师CATs使用手册》后、实际课上使用技术评价学生前调整其教学预设。教师也会在使用技术的过程中或刚刚完成技术时，迅速讲解并指导学生解决技术中的题目。遗憾的是，我们并没有发现教师使用技术揭示出的信息来调整后续教学以便满足学生的学习需求。如此，课堂评价技术并没有被作为评价活动使用，而是被用作教师教学计划中的补充习题。总的来说，教师认为课堂评价技术易于使用，能够揭示出学生数学学习的信息并把信息快速有效地呈献给教师，还能邀请学生广泛地参与到评价活动中来。

在第五章中，我们主要考察中国小学数学教师通过使用课堂评价技术能够洞悉到哪些有关学生数学理解的新信息。该研究的基本设计与上述探索性研究（见第四章）类似，但参与的教师人数更多，聚焦的数学内容也不同。具体地说，25位来自中国南京九所小学的三年级数学教师参与了这项研究。我们为他们提供了8个课堂评价技术用于评价学生对两位数乘法的理解。此外，我们还准备了《教师CATs使用手册》并召开了两次时长2小时的会议来帮助教师理解、使用这些课堂评价技术。教师在使用每个技术后需填写使用反馈表，章节教学结束后需完成课堂评价技术使用小结。当教师的回答满足下列两个条件时，我们就认为他们通过使用某个技术获得了对学生数学理解的新信息。第一个条件是教师指明该课堂评价技术所要评价的数学内容；第二个条件是教师提供下列三类信息中的至少一类：具体描述学生完成技术的表现，明确强调所获信息的新颖，或提出适切的教学改进。另外，为了探索“教师通过使用技术获得学生学习新信息”与“学生数学成绩变化”之间的关系，我们收集并分析了198名学生在参与研究前后的三次数学测试成绩。

在收集到的193份教师使用课堂评价技术的反馈中，有57份(30%)表明教师获得了对其学生数学学习情况的新认识。依据帮助教师洞悉新情况的技术数量，我们将参与该研究的教师分成三个组别。五位教师属**高洞察组**，因其在超过一半(五个或更多)的技术的使用过程中获取了新信息。十四位教师属**有洞察组**，他们在三个或更少的技术使用中发现新信息。其余六位教

师属**无洞察组**，他们没有通过使用技术得到任何新信息。课堂评价技术从与教师常规教学题目不同的角度来提问评价学生，因而具有揭示学生数学理解程度的能力。**高洞察组**的教师能够理解课堂评价技术的用意，并利用这些技术加深对学生的理解。而**无洞察组**的教师通常只概括、简单地指出其学生能否正确解题。他们中的一部分人认为课堂评价技术和教科书中的内容(非常)一致，所以不愿意重复已教授过的内容；另一部分则认为课堂评价技术与他们的常规教学相比太异太难。通过分析学生们的三次数学测试成绩得到，至少三分之二学生的成绩在85(共100分)以上。教师使用课堂评价技术后，**高洞察组**学生的成绩有小幅的提高，而**有洞察组**和**无洞察组**学生的成绩则小幅下降。然而，将标准化前测成绩作为协变量，洞察组别作为固定因子，进行协方差分析后，教师使用技术洞悉新信息的多少对学生成绩没有显著性影响。

在第六章中，我们总结了ICA-C研究项目中四个子课题的研究结果，并就小学数学教师的课堂评价实践和后续评价研究提出建议。ICA-C项目的研究结果表明：参与该课题的中国小学数学教师普遍认同运用评价来促进教和学这一理念。根据教师的反馈，他们在实践中使用各种方式评价学生对不同知识、技能的掌握情况以便实现多种评价目的，这与义务教育数学课程标准所倡导的相一致。但是，参与研究的教师似乎并没有利用评价信息来调整后续教学。课堂评价技术对于提高教师课堂评价活动是切实可行并有所帮助的，技术从有别于教科书习题的不同角度来评价学生，为教师深入了解学生的数学理解提供了新的途径。然而，使用课堂评价技术意味着教师要实施形成性评价，这对中国小学数学教师而言仍是一个挑战。想要充分利用课堂评价，可能需要给教师提供更多的帮助，而这需要政策制定者、教育研究者、教师培训者的共同努力。

Acknowledgements

On a sunny afternoon of my ordinary PhD working life, I came across a TEDx Talk given by Tracy McMillan, in which I found a sentence I must immediately write down: "*Life does not give you what you have asked for; it gives you the people, places, and situations that allow you to develop what you ask for*." In my life, I asked for becoming a researcher working with young children in mathematics, and then it brought me to the Netherlands to start my PhD adventure. In this section, I would like to thank the people who guided and supported me along this journey to help me become who I am today.

First of all, I would like to express my sincere and immense gratitude to my promotor, Marja van den Heuvel-Panhuizen. It is you who make my PhD adventure in the Netherlands possible by agreeing to supervise me. During these years, you have guided me with your impressive knowledge and expertise, critical thinking, tenacious work ethic, and true love and passion in mathematics education. This PhD thesis would have never been possible without all of your input and effort. I have benefited a lot from having an especially supportive promotor like you. You taught me by action how to do scientific research in the field of mathematics education, from setting up a study with a clear focus, to designing instructional material and assessment instruments, to collecting and analyzing data, and to writing scholarly articles with logic and precision. For each step, you reminded me of thinking critically. Also, you have created so many valuable opportunities for me to attend national and international conferences to further broaden my knowledge and experience. Thanks to you, I deepen my understanding of classroom assessment and develop the skills which are essential for my future academic career.

I would like to extend my heartfelt and special thanks also to Michiel Veldhuis, my copromotor. You have witnessed all my changes and growth on the way of doing a PhD. Over these years, you have asked me many good and critical questions that make me thoroughly reflect on the research, you have provided me lots of constructive and actionable advices to deal with the specific issues or difficulties that I encountered, you have explained to me the important knowledge and skills, statistical analyses in particular, for doing the research, and you set a good example for me regarding communicating precisely and effectively in the team work. You were also very important for finishing the thesis. Besides the professional support, thank you for all the great encouragement and help in my personal life.

I want to acknowledge the China Scholarship Council (CSC) for the scholarship I received. Specifically, I would like to express my appreciation to Counselor Qingyu Meng, Yiwei Wang, Lei Xia, and Xiaoxiao Chen for your understanding and encouragement which helped me fulfill my PhD study.

During my PhD study I received a lot of supports from my friends and colleagues I met at the Freudenthal Institute and the Freudenthal Group. Nathalie Martel and Ariyadi Wijaya, I was so lucky to become your officemate. Thank you for all the experience and knowledge you shared with me, which helped me a lot to start and continue my work and life in the Netherlands. Adri Dierdorp, thank you for your kindness and sharing, I enjoyed every talk we had. Al Jupri, Ngô Vũ Thu Hăng, Shinya Itoh, Jutta Reuwsaat, and Ednei Becher, thank you for bring me different perspectives of culture regarding mathematics education as well as daily work and life. Michiel Doorman, I am very grateful for all your lectures and examples which showed me how to observe life with the eyes of mathematics. Thanks to you, I started to ask myself quite often whether and what mathematics I can see from here. Also, thanks for the good memory during the visit in China. Martin Kindt, I admire you for your expertise and passion in teaching algebra. I have learned a lot from your interesting PPT slides and our informative discussion about the cake problem. Annemiek van Leendert, thank you for supporting me in this PhD journey, I felt much more motivated every time after our meeting. Mark uwland, Liesbeth Walther, and Liesbeth der Bakker, thank you for the help in daily work and your kindness and attention that made me feel to be part of the social events. Sylvia van Borkulo, Peter Boon, Sietske Tocoma, and Dolly van Eerde, I am grateful for all the encouraging and inspiring conversations we had. Nathalie Kuijpers, many thanks for all the English checking, the editorial assistance, and making the beautiful thesis cover.

As a PhD student, life sometimes can be very difficult. Ilona Friso-van den Bos, you are one of my great sources of inspiration, encouragement and comfort. I appreciate all of your careful and patient listening, informative opinions and constructive advices. Also, I am lucky to have wonderful companions in this

PhD journey, with whom I shared the same goal. Ali Sangari, thank you for your sense of humor, which lights up our daily working. I have learned a lot from you about courage, perseverance, and dedication. Many thanks for all the kindness, sharing, and considerations I got from both you and Reyhane. Mara Otten, Roos Blankespoor, and Carolien Duijzer, you are the biggest surprise in my PhD days, in the way like how you showed up in the dance event. I cannot describe how cheerful and enjoyable to work with you. I learned a lot from you to be independent, creative and strong. Mara Otten, I deeply appreciate you for taking care of me, sharing with me your opinions and feelings, and creating me a good memory of playing football in the Netherlands. Roos Blankespoor, I am very grateful for all the inspiring and enjoyable discussions we had. I will remember all the metaphors you told me forever. Carolien Duijzer, I guess you have a special toolkit which is full of great and cool ideas. Thanks a lot for all the sharing from which both Jiakun and I benefited.

Lots of the work in this PhD research has been done in Nanjing, China, where I received great help from my previous teachers and many schoolfellows in Nanjing Normal University. Rongbao Tu, your introduction of Realistic Mathematics Education (RME) and Hans Freudenthal opened my eyes, and your trust and encouragement made me start to imagine studying in Freudenthal Institute in the Netherlands. Lianhua Ning, from the very beginning when I applied for the CSC scholarship, you have already offered great support. During my PhD journey, you continued helping to download hundreds of Chinese literatures and to contact the schools and the teachers for participation in three studies. Thanks to you, all the data collection in China went smoothly. I also would like to thank Guangming Wei, Linhong Hao, and Yucheng Zhou, together with the teachers and the students in your schools, for your interests in using classroom assessment techniques and your participation and contribution in the studies. Suhong Zhang, Qun Hui, Lihua Cui, Jie Wang, Yanjing Guo, Hai Huang, Li Wu, Tiantian Zhang, Ningning Zhang, Yang Shen, Liqian Xie, Wei Zhang, Dawei Li, Gan Pan and Jun Dong, I am very grateful for receiving your help in several rounds of the data collection.

Besides working on the PhD research, I appreciate to have the opportunities to attend conferences held in China and worldwide. Special thanks go to Xiaotian Sun and Wei He from Minzu University of China for organizing the forums on RME. You have created valuable opportunities for me to become a part of the

forums, which not only deepened my understanding of RME but also updated my knowledge of mathematics education in mainland China. Fang Liang, Xujie Jia, Ao Xue, Jing Wang and Lianchun Dong, thank you for showing me how to work as a team productively and efficiently. It was such a pleasure to work with you. Also, I want to thank all the teachers from New Century Nine Chapter Technology Educational Institute in Beijing and ZheJiang New Ideas Institute of Educational Sciences in Hangzhou. I have learned a lot from working with you. For colleagues I met from ICO, ICMI Study 23, ICME-13, and CERME 10, thank you for the inspiring and informative discussions from which I broadened my knowledge and skills of doing research in mathematics education and became more motivated to finishing my PhD study.

It is hard to study and live in a foreign country, especially considering that I had never been to any other contries before coming to the Netherlands. I would like take this opportunity to thank the people who supported me in my Dutch adventure. Janice Rossen and Bill Rossen, many thanks for your warm welcome when I arrived here, which means a lot to me. Over these years, you have been always taking me into account. And you have provided great help and encouragement to both Jiakun and me in many crucial moments when we encountered difficulties, not only in academic field but also in daily life. Hein van Vliet, I am very grateful for your kindness, generosity and openness. In our meetings, you helped me a lot to improve my English. Thanks to your explanations and suggestions, I became gradually familiar to the Dutch society, which indeed made me live here in a better way. Cristina Morais, I am really thankful that life brings you to me. If I look back now, it was amazing how we started to talk, to know, and to care about each other. I admire you a lot for your passion, energy, and dedication to do research in mathematics education. And I enjoyed so much for every contact and talk we had either via Internet or face to face in a conference. We definitely need to do it more often.

Only when I am far away from my motherland does it become clear to me how important for me being part of the Chinese community. Tianqi Li and Xuexue Chen, I heard a lot about you even before I came to the Netherlands. Thanks for all the help you offered to me. Lizuo Xin and Yanbo Wang, Ping Liu and Qingbao Guan, Feifei Song and Xinyuan Mao, Ying Yang and Xuliang Hou, Jiao Chen and Fei Xu, Zhaokun Guo and Xiang Fu, Haoyu Wang, Jing Lv, and Ding Ding, thanks to all of you, my days in Delft become meaningful and colorful. I learned a lot from you about how to care and support each other in front of various challenges in both academic world and daily life. Lijie Zheng, Honghong Bai, Yuru Wang, Mei Liu and Xiaojing Sun, I am very glad that we had more opportunity to spend time together after I moved to Langeveld Building. As all of our research has to do with education in one way or another, I gained a lot from you about different educational studies. My special thanks go to Hongbo Guo. A simple "Hi" on the way cycling to Uithof made the start of our friendship. There were so many days we spoke out the hopes in the morning and share the working details in the evening. I feel so lucky travelling together with you, not only between home and office every day, but also on the path to pursue the doctors' degree. Eventually, we, as well as Wenshi Wang, make it in 2018!

Also, I want to extend my appreciation to my old friends: Manman Chen, Fan Yang, Fangfang Lv, Huafang Yang, Jingjing Wu, and Lanjie Sun. Thank you for all the stories you shared with me from far away. It is hard to describe how powerful and encouraging those conversations could be to me.

Finally, I owe a debt of gratitude to my families. My parents-in-law and my parents, without your unconditional love, understanding and support, I would never go abroad for this PhD study and would never make it in the end. 亲爱的爸妈们，希望此刻女儿足够让您骄傲！To all of my relatives, thank you for the support. Every time when you told me that "Don't worry about the things at home; we are all here", I felt deeply moved. Jiakun, my dear husband, it is you who planted the seed of studying abroad in my mind, it is you who has witnessed every single step I took in this journey, and it is you who are still standing firmly behind me, as always. "静心，尽力，结果寓于过程之中", literally means keeping a peaceful heart and making every effort in the process and then the outcome has already been implied, with this wisdom you told me, I am ready to set out, together with you, to explore the world in the next stage of our life.

Curriculum vitae

Xiaoyan Zhao was born on November 01, 1985 in Zhangjiakou, Hebei, China. For secondary education, she studied at Xuanhua No. 1 Secondary School from 2001 to 2004 and at Innovation International Secondary School in Zhangjiakou for the year next. In 2005, she enrolled in a bachelor program in Mathematics and Applied Mathematics at Hebei Normal University, and obtained her bachelor's degree in Science in 2009. In the same year, she began to study in Nanjing Normal University and got her master's degree in Education in 2012. In September 2012, Xiaoyan Zhao received a scholarship from the China Scholarship Council, and started her PhD research at the Freudenthal Institute in Utrecht University in the Netherlands. The PhD research was carried out under the supervision of Prof. dr. Marja van den Heuvel-Panhuizen and Dr. Michiel Veldhuis, with focus on Chinese primary school mathematics teachers' classroom assessment. The research findings have been published in several international journal articles, and have been presented at national and international conferences. Also, Xiaoyan Zhao fulfilled all requirements of the Interuniversity Center for Educational Sciences (ICO) Research School in the Netherlands as a PhD member. Recently, Xiaoyan is going back to China, and she will continue her research in mathematics education in primary school.

List of publications related to this thesis

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. (2017). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education,* Advance online publication, 1–21.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2017). Using classroom assessment techniques in Chinese primary schools: Effects on student achievement. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (pp. 3604–3611). Dublin, Ireland: DCU Institute of Education and ERME.

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), 1–18.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2015). Classroom assessment techniques to assess Chinese students' sense of division. In X. Sun, B. Kaur, & J. Novotná (Eds.), *Conference Proceedings of ICMI Study 23: Primary mathematics study on whole numbers* (pp. 496–503). Macau, China: University of Macau.

Article currently (January 2018) under review:

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2017). *Insights Chinese primary mathematics teachers gained into their students' learning from using classroom assessment techniques*. Submitted for review.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2018). *Assessment in the hands of Chinese primary school mathematics teachers.* Submitted as a PME Research Report.

List of presentations related to this thesis

Zhao, X. (2017). *Mathematics education in two worlds: A personal reflection*. Presentation at the Third Forum on Realistic Mathematics Education & Eighth Elementary Mathematics Education Summit in China, Hangzhou, China, September 25–27.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2017). *Supporting Chinese primary school teachers' assessment practice in mathematics: effects on student achievement*. Paper presented at the the Tenth Congress of the European Society for Research in Mathematics Education, Dublin, Ireland, February 1–5.

Veldhuis, M., Van den Heuvel-Panhuizen, M. & Zhao, X. (2016). *Soutien aux enseignants dans leur pratique d'évaluation en mathématiques : effets sur les compétences des élèves.* Paper presented at Colloque International Évaluation en Mathématiques, Paris, France, November 21–22.

Veldhuis, M., Van den Heuvel-Panhuizen, M. & Zhao, X. (2016). *Supporting primary teachers' classroom assessment in mathematics: effects on student learning in the Netherlands and Nanjing, China*. Paper presented at the Thirteenth International Congress on Mathematics Education, Hamburg, Germany, July 24–31.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2015). *A survey of teachers' classroom assessment practice in primary mathematics education in China*. Poster presented at the ICO National Fall School, Utrecht, the Netherlands, November 5–6.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2015). *Chinese teachers' use of classroom assessment techniques in primary mathematics education*. Paper presented at the Onderwijs Research Dagen 2015, Leiden, the Netherlands, June 17–19.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2015). *Classroom assessment techniques to assess Chinese students' sense of division*. Paper presented at the Twenty-third ICMI Study: Primary Mathematics Study on Whole Numbers, Macao, China, June 3–7.

Zhao, X., Van den Heuvel-Panhuizen, M. & Veldhuis, M. (2014). *Chinese Teachers' First Experience with Classroom Assessment Techniques (CATs) in Primary Mathematics Education*. Paper presented at the ICO International Fall School, Blankenberge, Belguim, November 10–14.

FI Scientific Library

(formerly published as FIsme Scientific Library, CD-β Scientific Library)

95. Van der Laan, S. (2017). *Een varken voor iedereen – de modernisering van de Nederlandse varkensfokkerij in de twintigste eeuw.*
94. Vis, C. (2017). *Strengthening local curricular capacity in international development cooperation.*
93. Benedictus, F. J. (2017). *Reichenbach: Probability and the A Priori. Has The Baby Been Thrown Out with the Bathwater?*
92. De Ruiter, P. A. C. (2016). *Het Mijnwezen in Nederlands-Oost-Indië 1850-1950*.
91. Roersch van der Hoogte, A. (2015). *Colonial Agro-Industrialism. Science, industry and the state in the Dutch Golden Alkaloid Age, 1850-1950*.
90. Veldhuis, M. (2015). *Improving classroom assessment in primary mathematics education*.
89. Jupri, A. (2015). *The use of applets to improve Indonesian student performance in algebra*.
88. Wijaya, A. (2015). *Context-based mathematics tasks in Indonesia: Toward better practice and achievement*.
87. Klerk, S. (2015). *Galen reconsidered. Studying drug properties and the foundations of medicine in the Dutch Republic ca. 1550-1700*.
86. Krüger, J. H. J. (2014). *Actoren en factoren achter het wiskundecurriculum sinds 1600.*
85. Lijnse, P. L. (2014). *Omzien in verwondering. Een persoonlijke terugblik op 40 jaar werken in de natuurkundedidactiek*.
84. Weelie, D. van (2014). *Recontextualiseren van het concept biodiversiteit.*
83. Bakker, M. (2014). *Using mini-games for learning multiplication and division: a longitudinal effect study*.
82. Ngô Vũ Thu Hằng (2014). *Design of a social constructivism-based curriculum for primary science education in Confucian heritage culture*.
81. Sun, L. (2014). *From rhetoric to practice: enhancing environmental literacy of pupils in China*.
80. Mazereeuw, M. (2013). *The functionality of biological knowledge in the workplace. Integrating school and workplace learning about reproduction.*
79. Dierdorp, A. (2013). *Learning correlation and regression within authentic contexts.*
78. Dolfing, R. (2013). *Teachers' Professional Development in Context-based Chemistry Education. Strategies to Support Teachers in Developing Domainspecific Expertise.*
77. Mil, M. H. W. van (2013). *Learning and teaching the molecular basis of life.*

76. Antwi, V. (2013). *Interactive teaching of mechanics in a Ghanaian university context.*
75. Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms.*
74. Stolk, M. J. (2013). *Empowering chemistry teachers for context-based education. Towards a framework for design and evaluation of a teacher professional development programme in curriculum innovations.*
73. Agung, S. (2013). *Facilitating professional development of Madrasah chemistry teachers. Analysis of its establishment in the decentralized educational system of Indonesia.*
72. Wierdsma, M. (2012). *Recontextualising cellular respiration.*
71. Peltenburg, M. (2012). *Mathematical potential of special education students.*
70. Moolenbroek, A. van (2012). *Be aware of behaviour. Learning and teaching behavioural biology in secondary education.*
69. Prins, G. T., Vos, M. A. J., & Pilot, A. (2011). *Leerlingpercepties van onderzoek & ontwerpen in het technasium.*
68. Bokhove, Chr. (2011). *Use of ICT for acquiring, practicing and assessing algebraic expertise.*
67. Boerwinkel, D. J. & Waarlo, A. J. (2011). *Genomics education for decisionmaking. Proceedings of the second invitational workshop on genomics education, 2-3 December 2010.*
66. Kolovou, A. (2011). *Mathematical problem solving in primary school*.
65. Meijer, M. R. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study*.
64. Kortland, J., & Klaassen, C. J. W. M. (2010). *Designing theory-based teaching-learning sequences for science. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of Physics Didactics at Utrecht University*.
63. Prins, G. T. (2010).*Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning*.
62. Boerwinkel, D. J., & Waarlo, A. J. (2010). *Rethinking science curricula in the genomics era. Proceedings of an invitational workshop*.
61. Ormel, B. J. B. (2010). *Het natuurwetenschappelijk modelleren van dynamische systemen. Naar een didactiek voor het voortgezet onderwijs*.
60. Hammann, M., Waarlo, A. J., & Boersma, K. Th. (Eds.) (2010). *The nature of research in biological education: Old and new perspectives on theoretical and methodological issues – A selection of papers presented at the VIIth Conference of European Researchers in Didactics of Biology*.
59. Van Nes, F. (2009). *Young children's spatial structuring ability and emerging number sense*.

58. Engelbarts, M. (2009). *Op weg naar een didactiek voor natuurkunde-experimenten op afstand. Ontwerp en evaluatie van een via internet uitvoerbaar experiment voor leerlingen uit het voortgezet onderwijs*.
57. Buijs, K. (2008). *Leren vermenigvuldigen met meercijferige getallen*.
56. Westra, R. H. V. (2008). *Learning and teaching ecosystem behaviour in secondary education: Systems thinking and modelling in authentic practices*.
55. Hovinga, D. (2007). *Ont-dekken en toe-dekken: Leren over de veelvormige relatie van mensen met natuur in NME-leertrajecten duurzame ontwikkeling.*
54. Westra, A. S. (2006). *A new approach to teaching and learning mechanics*.
53. Van Berkel, B. (2005). *The structure of school chemistry: A quest for conditions for escape.*
52. Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education: The case of water quality*.
51. Doorman, L. M. (2005). *Modelling motion: from trace graphs to instantaneous change.*
50. Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools.*
49. Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education*.
48. Drijvers, P. (2003). *Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter.*
47. Van den Boer, C. (2003). *Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs.*
46. Boerwinkel, D. J. (2003). *Het vormfunctieperspectief als leerdoel van natuuronderwijs. Leren kijken door de ontwerpersbril*.
45. Keijzer, R. (2003). *Teaching formal mathematics in primary education. Fraction learning as mathematising process.*
44. Smits, Th. J. M. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten.*
43. Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education – The yo-yo learning and teaching strategy*.
42. Dressler, M. (2002). *Education in Israel on collaborative management of shared water resources.*
41. Van Amerom, B.A. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*.
40. Van Groenestijn, M. (2002). *A gateway to numeracy. A study of numeracy in adult basic education.*

39. Menne, J. J. M. (2001). *Met sprongen vooruit: een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment.*
38. De Jong, O., Savelsbergh, E.R., & Alblas, A. (2001). *Teaching for scientific literacy: context, competency, and curriculum.*
37. Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue.*
36. Lijmbach, S., Broens, M., & Hovinga, D. (2000). *Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking.*
35. Margadant-van Arcken, M., & Van den Berg, C. (2000). *Natuur in pluralistisch perspectief – Theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden.*
34. *Janssen, F. J. J. M. (1999). Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs.*
33. De Moor, E. W. A. (1999). *Van vormleer naar realistische meetkunde – Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw.*
32. Van den Heuvel-Panhuizen, M., & Vermeer, H. J. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek.*
31. Beeftink, C. (2000). *Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs.*
30. Vollebregt, M. J. (1998). *A problem posing approach to teaching an initial particle model.*
29. Klein, A. S. (1998). *Flexibilization of mental arithmeticsstrategies on a different knowledge base – The empty number line in a realistic versus gradual program design.*
28. Genseberger, R. (1997). *Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer.*
27. Kaper, W. H. (1997). *Thermodynamica leren onderwijzen*.
26. Gravemeijer, K. (1997). *The role of context and models in the development of mathematical strategies and procedures*.
25. Acampo, J. J. C. (1997). *Teaching electrochemical cells – A study on teachers' conceptions and teaching problems in secondary education*.
24. Reygel, P. C. F. (1997). *Het thema 'reproductie' in het schoolvak biologie*.
23. Roebertsen, H. (1996). *Integratie en toepassing van biologische kennis – Ontwikkeling en onderzoek van een curriculum rond het thema 'Lichaamsprocessen en Vergift'*.

22. Lijnse, P. L., & Wubbels, T. (1996). *Over natuurkundedidactiek, curriculumontwikkeling en lerarenopleiding*.
21. Buddingh', J. (1997). *Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek*.
20. Van Hoeve-Brouwer G. M. (1996). *Teaching structures in chemistry – An educational structure for chemical bonding*.
19. Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education.*
18. Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity.*
17. De Jong, O., Van Roon, P. H., & De Vos, W. (1995). *Perspectives on research in chemical education.*
16. Van Keulen, H. (1995). *Making sense – Simulation-of-research in organic chemistry education.*
15. Doorman, L. M., Drijvers, P. & Kindt, M. (1994). *De grafische rekenmachine in het wiskundeonderwijs.*
14. Gravemeijer, K. (1994). *Realistic mathematics education.*
13. Lijnse, P. L. (Ed.) (1993). *European research in science education.*
12. Zuidema, J., & Van der Gaag, L. (1993). *De volgende opgave van de computer.*
11. Gravemeijer, K., Van den Heuvel-Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek.*
10. Van der Valk, A. E. (1992). *Ontwikkeling in Energieonderwijs*.
9. Streefland, L. (Ed.) (1991). *Realistic mathematics education in primary schools.*
8. Van Galen, F., Dolk, M., Feijs, E., & Jonker, V. (1991). *Interactieve video in de nascholing reken-wiskunde.*
7. Elzenga, H. E. (1991). *Kwaliteit van kwantiteit.*
6. Lijnse, P. L., Licht, P., De Vos, W., & Waarlo, A. J. (Eds.) (1990). *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education.*
5. Van Driel, J. H. (1990). *Betrokken bij evenwicht.*
4. Vogelezang, M. J. (1990). *Een onverdeelbare eenheid.*
3. Wierstra, R. F. A. (1990). *Natuurkunde-onderwijs tussen leefwereld en vakstructuur.*
2. Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education.*
1. Lijnse, P. L., & De Vos, W. (Eds.) (1990). *Didactiek in perspectief.*

ICO Dissertation Series

In the ICO Dissertation Series, dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University, and formerly University of Groningen (until 2006), Radboud University Nijmegen (until 2004), and Tilburg University (until 2002). The University of Groningen, University of Antwerp, University of Ghent, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

Recent publications in the ICO Dissertation Series (updated January, 2018):

372. Van Dijk, A. M. (22-12-2017). *Learning together in mixed-ability elementary classrooms.* Enschede: University of Twente.
371. Rezende Da Cunha Junior, F. (19-12-2017). *Online groups in secondary education.* Amsterdam: Vrije Universiteit Amsterdam.
370. Frèrejean, J. (08-12-2017). *Instruction for information problem solving.* Heerlen: Open University of the Netherlands.
369. Van der Veen, M. (28-11-2017). *Dialogic classroom talk in early childhood education*. Amsterdam: Vrije Universiteit Amsterdam.
368. Stoel, G. L. (14-11-2017). *Teaching towards historical expertise. Developing students' ability to reason causally in history.* Amsterdam: University of Amsterdam.
367. Binkhorst, F. (20-10-2017). *Connecting the dots. Supporting the implementation of Teacher Design Teams.* Enschede: University of Twente.
366. Wansink, B. G. J. (15-09-2017). *Between fact and interpretation. Teachers' beliefs and practices in interpretational history teaching.* Utrecht: Utrecht University.
365. Moses, I. (07-09-2017). *Student-teachers' commitment to teaching*. Leiden: Leiden University.
364. Evens, M. (30-06-2017). *Pedagogical content knowledge of French as a foreign language: Unraveling its development*. Leuven: KU Leuven.

363. Van Waes, S. (23-06-2017). *The ties that teach: Teaching networks in higher education.* Antwerp: University of Antwerp.
362. Tuithof, J. I. G. M. (23-06-2017). *The characteristics of Dutch experienced history teachers' PCK in the context of a curriculum innovation.* Utrecht: Utrecht University.
361. Janssen, N. (23-06-2017). *Supporting teachers' technology integration in lesson plans.* Enschede: University of Twente.
360. Van Lankveld, T. A. M. (20-06-2017). *Strengthening medical teachers' professional identity. Understanding identity development and the role of teacher communities and teaching courses*. Amsterdam: Vrije Universiteit Amsterdam.
359. Brouwer, J. (22-05-2017). *Connecting, interacting and supporting. Social capital, peer network and cognitive perspectives on small group teaching*. Groningen: University of Groningen.
358. Grohnert, T. (18-05-2017). *Judge/Fail/Learn; enabling auditors to make high-quality judgments by designing effective learning environments.* Maastricht: Maastricht University.
357. Lans, R. M. (18-05-2017). *Teacher evaluation through observation.* Groningen: University of Groningen.
356. Pennings, J. M. (04-05-2017). *Interpersonal dynamics in teacher-student interactions and relationships.* Utrecht: Utrecht University.
355. Baas, D. M. (01-05-2017). *Assessment for Learning: more than a tool.* Maastricht: Maastricht University.
354. Wouters, A. (09-02-2017). *Effects of medical school selection. On the motivation of the student population and applicant pool.* Amsterdam: VU Medisch Centrum.
353. Baggen, Y. (13-01-2017). *LLLIGHT 'in' Europe - Lifelong Learning, Innovation, Growth and Human capital Tracks in Europe*. Wageningen: Wageningen University.
352. Osagie, E. R. (14-12-2016) *Learning and Corporate Social Responsibility. A study on the role of the learning organization, individual competencies, goal orientation and the learning climate in the CSR adaptation process.* Wageningen: Wageningen University.
351. Beckers, J. (09-12-2016). *With a little help from my e-portfolio; supporting students' self directed learning in senior vocational education.* Maastricht: Maastricht University.
350. Oonk, C. (07-12-2016). *Learning and Teaching in the Regional Learning Environment: Enabling Students and Teachers to Cross Boundaries in Multi-Stakeholder Practices'*. Wageningen: Wageningen University

Universiteit Utrecht

In mainland China, where there exists a deeply-rooted examination culture, an assessment reform promoting the use of assessment to support teaching and learning has been carried out since 2001. After a decade, however, only a few studies have been done that focus on primary school mathematics teachers' assessment practice at the classroom level. With this background, the Improving Classroom Assessment in China (ICA-C) project, as a sequel to the ICA project in the Netherlands, was set up to investigate the current situation of Chinese primary school mathematics teachers' perceptions and practices of classroom assessment, and to explore the possibility for improving their assessment activities by using classroom assessment techniques (CATs). CATs are short teacher-initiated targeted assessment activities proximate to the textbook, which teachers can use in their daily practice to make informed instructional decisions. The results of the ICA-C project show that the involved teachers generally endorsed the idea of using assessment to improve teaching and learning. CATs are feasible and useful to enhance teachers' classroom assessment activities. Nevertheless, using assessment information for adapting further instruction is challenging for Chinese primary school mathematics teachers. More support is necessary for helping them make the most of their classroom assessment.