Zhao et al. International Journal of STEM Education (2016) 3:19
DOI 10.1186/s40594-016-0051-2

International Journal of STEM Education

RESEARCH Open Access

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

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

Abstract

Background: 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.

Results: 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.

Conclusions: 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

* Correspondence: x.zhao@uu.nl
[1]Freudenthal Institute, Faculty of Science, Utrecht University, PO Box 85170, Utrecht 3508 AD, the Netherlands
Full list of author information is available at the end of the article

Springer Open

© The Author(s). 2016 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

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.

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 and 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 and James 2014; Hill and McNamara 2012).

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 et al. (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 and Van den Heuvel-Panhuizen 2014; 2016). 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 and 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 and Van den Heuvel-Panhuizen 2016).

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 and 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 et al. 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 et al. (2016a) 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 et al. 2016b) on teachers' assessment practice and beliefs in primary mathematics classes in China, it was revealed, based on 1158 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.

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 et al. (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 and Van den Heuvel-Panhuizen 2014; 2016; 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 and different classroom practices (see, 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 et al. 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?

Methods

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.

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.

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 et al. 2009) and pay much attention to study and understand textbooks carefully and thoroughly (Cai and Wang 2010). In this way, the CATs could be embedded in the teachers' daily classroom practice. The six teachers all used the 苏教版 Textbook, published by Jiangsu Education Publishing House (2005).

Based on the characteristics of CATs developed in the Netherlands (Veldhuis and 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.

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 苏教版 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.

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 (cf. the "Content" column in Table 1).

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 (see 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 (see Fig. 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 (see Fig. 2).

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 (see 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.

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-

Table 1 Multiplication and division in苏教版 Textbook (Grade 2 – Grade 4)

Grade-Semester	Chapter	Topic	Content	Example problem		
				Horizontal notation		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			$\begin{array}{r} 4 \\ \times\ 2 \\ \hline 8 \end{array}$ $\begin{array}{r} 3 \\ 2\overline{)6} \\ \underline{6} \\ 0 \end{array}$
2-2	1	Division	Division with remainder; 1- or 2-digit divided by 1-digit	7÷3	17÷5	$\begin{array}{r} 2 \\ 3\overline{)7} \\ \underline{6} \\ 1 \end{array}$ $\begin{array}{r} 3 \\ 5\overline{)17} \\ \underline{15} \\ 2 \end{array}$
	8	Multiplication	2-digit multiplied by 1-digit	20×3		
				14×2	48×2	$\begin{array}{r} 14 \\ \times\ \ 2 \\ \hline 28 \end{array}$ $\begin{array}{r} 48 \\ \times\ \ 2 \\ \hline 96 \end{array}$
3-1	1	Division	2-digit divided by 1-digit	40÷2		$\begin{array}{r} 26 \\ 2\overline{)52} \\ \underline{4} \\ 12 \\ \underline{12} \\ 0 \end{array}$ $\begin{array}{r} 20 \\ 3\overline{)62} \\ \underline{6} \\ 2 \end{array}$
				52÷2	62÷3	
	7	Multiplication	3-digit multiplied by 1-digit	400×2		
				120×4		$\begin{array}{r} 152 \\ \times\ \ 4 \\ \hline 608 \end{array}$ $\begin{array}{r} 120 \\ \times\ \ 4 \\ \hline 480 \end{array}$
3-2	1	Division	3-digit divided by 1-digit	600÷3		
				306÷3		$\begin{array}{r} 78 \\ 4\overline{)312} \\ \underline{28} \\ 32 \\ \underline{32} \\ 0 \end{array}$ $\begin{array}{r} 102 \\ 3\overline{)306} \\ \underline{3} \\ 6 \\ \underline{6} \\ 0 \end{array}$
	4	Multiplication	2-digit multiplied by 2-digit	12×10		
				25×30		$\begin{array}{r} 28 \\ \times\ 12 \\ \hline 56 \\ \underline{28} \\ 336 \end{array}$ $\begin{array}{r} 25 \\ \times\ 30 \\ \hline 750 \end{array}$
4-1	1	Division	2- or 3-digit divided by 2-digit	60÷20		$\begin{array}{r} 3 \\ 20\overline{)60} \\ \underline{60} \\ 0 \end{array}$ $\begin{array}{r} 8 \\ 34\overline{)272} \\ \underline{272} \\ 0 \end{array}$
4-2	1	Multiplication	3-digit multiplied by 2-digit			$\begin{array}{r} 144 \\ \times\ \ 15 \\ \hline 720 \\ \underline{144} \\ 2160 \end{array}$ $\begin{array}{r} 850 \\ \times\ \ 20 \\ \hline 17000 \end{array}$

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 (see 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

Table 2 Lessons in Chapter 1: division of 3-digit numbers by 1-digit numbers

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)			

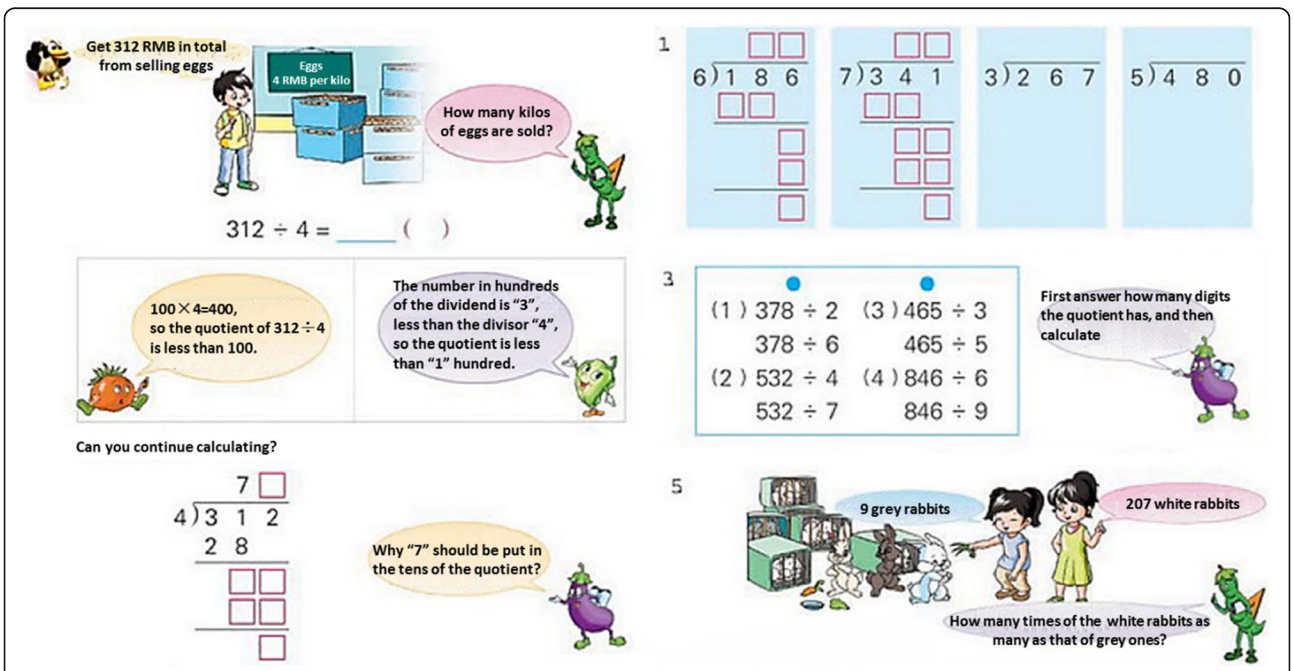


Fig. 1 Example problem in lesson 2 (*left*) and some corresponding exercises (*right*) (translated from Chinese)

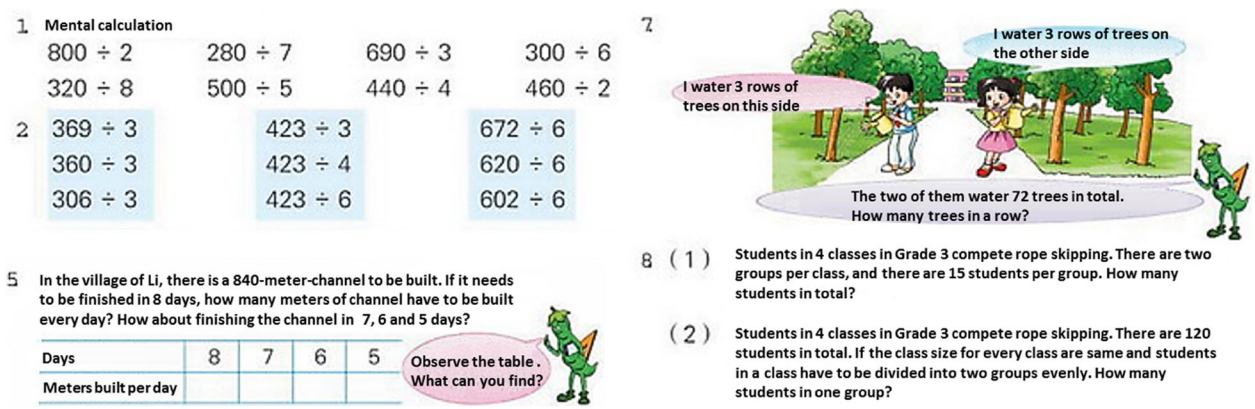


Fig. 2 Part of the exercises in lesson 8 (translated from Chinese)

quotients of two digits, like in $312 \div 4$. 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 (see Fig. 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 (see Fig. 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 whole-classroom 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

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

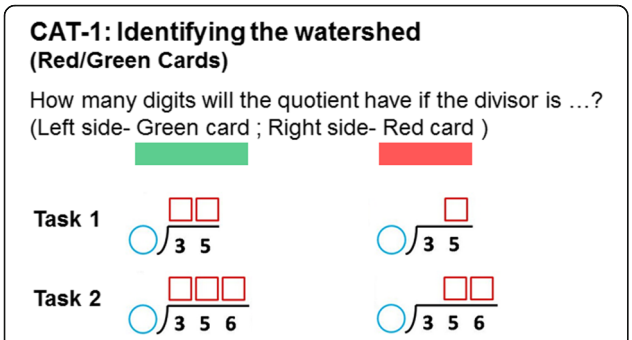
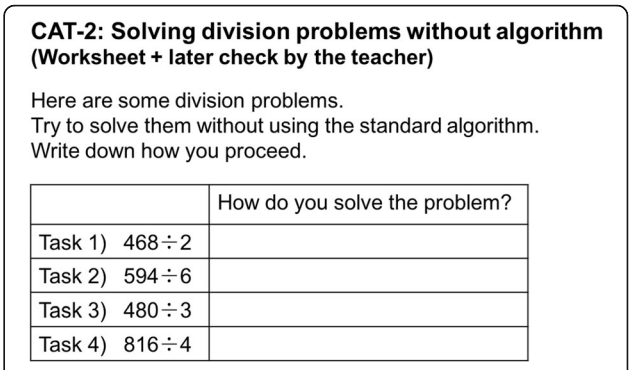


Fig. 3 CAT-1: identifying the watershed (translated from Chinese)



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	

Fig. 4 CAT-2: solving division problems without algorithm (translated from Chinese)

vary the content of these tasks and the number of tasks they use.

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.

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 (see Fig. 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.

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.

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-h 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 and 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 (cf. 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.

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.

Results

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.

Results for CAT-1: identifying the watershed

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 (see Fig. 5) and emphasized the divisors corresponding to the green card by drawing an accolade.

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

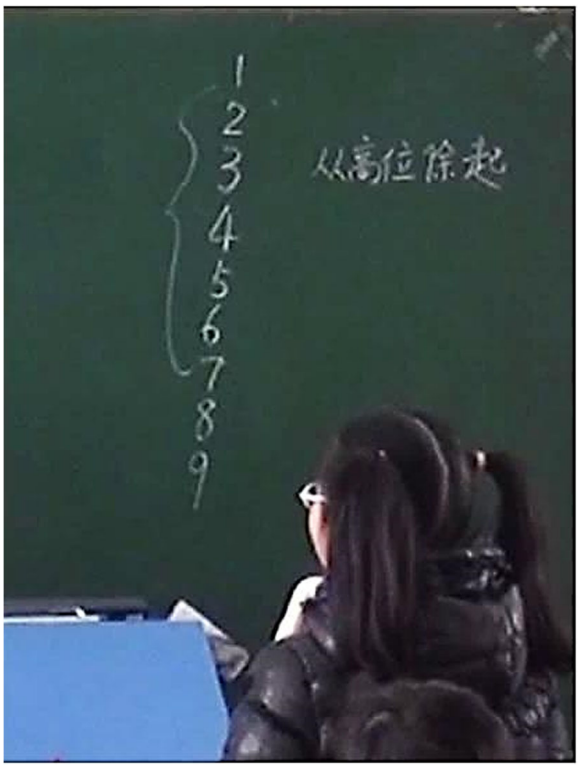


Fig. 5 Teacher C's use of CAT-1 showing the watershed with an accolade for dividend as 721

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.

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:

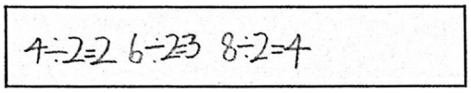
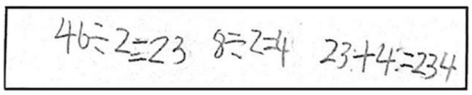
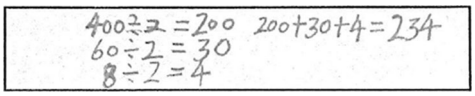
> In general, the exercises given to students ultimately focus on calculation, even if students were asked to make a decision about how many digits the quotient has [see Fig. 1, exercise 3]. But if [students were] not [asked to] calculate, more attention will be paid to understanding. (Teacher A in interview;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

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 %	 
b	Whole-number-based splitting the dividend	40 %	

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

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.

Results for CAT-2: solving division problems without algorithm

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.

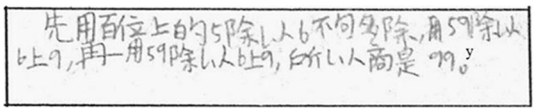
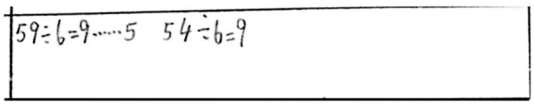
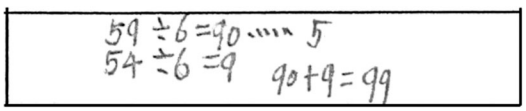
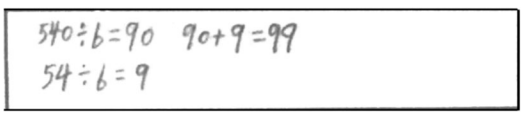
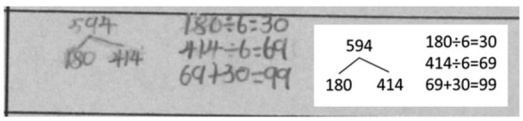
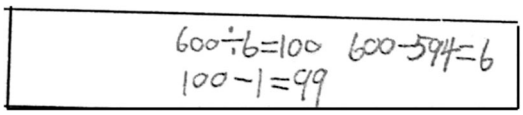
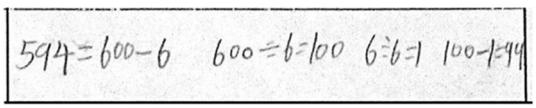
Our analysis of the student work of CAT-2

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).

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 (see Table 5 (a)) or notating the algorithm in a horizontal digit-based way (see 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 (see 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 (see 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 (see 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 %	
b	Digit-based splitting the dividend	65 %	
c	Digit-based splitting the dividend with the answer expressed as a whole number	10 %	
d	Whole-number-based splitting the dividend	12 %	
			
e	Whole-number-based changing the dividend	7 %	
			

[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 ($N = 127$)

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 (see 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 (see 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 (see 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.

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)

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 and 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 et al. (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.

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 self-evident 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.

Abbreviations
CATs: Classroom assessment techniques; MoE: Ministry of 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. 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.

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.

Author details
[1]Freudenthal Institute, Faculty of Science, Utrecht University, PO Box 85170, Utrecht 3508 AD, the Netherlands. [2]Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, PO Box 80140, Utrecht 3508 TC, the Netherlands.

Received: 25 April 2016 Accepted: 21 September 2016
Published online: 05 November 2016

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) "集体备课"辨析 [An analysis of the practice of "collective preparation of lessons"]. 中国教育学刊, *9*, 40-41.

Cui, Y. (2008). 教师应先学会评价再学习上课 [Teachers need to learn how to assess before to learn how to teach]. 基础教育课程, *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). 苏教版教科书(小学数学三年级下册) *[苏教版Textbook (Mathematics textbook for Grade 3 in primary education, Volume 2)]*. Nanjing: Author.

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). "集体备课": 内涵、问题与变革策略 [Collective lesson planning: its meaning, problems and strategies for change]. 西北师大学报*(社会科学版), 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 – The International Journal on Mathematics Education, 41*(6), 717–731.

Ministry of Education of the People's Republic of China (MoE). (2001). 基础教育课程改革纲要(试行) *[The Curriculum reform guidelines for the nine-year compulsory education (trial version)]*. Retrieved from http://moe.edu.cn/publicfiles/business/htmlfiles/moe/s8001/201404/167343.html. Accessed 19 June 2014.

Ministry of Education of the People's Republic of China (MoE). (2011). 义务教育数学课程标准(2011年版) *[Mathematics curriculum standards for nine-year compulsory education (2011 version)]*. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s8001/201404/167340.html. Accessed 19 June 2014.

OECD. (2012). *Education at a Glance 2012: OECD Indicators*. http://www.oecd.org/edu/EAG%202012_e-book_EN_200912.pdf. Accessed 19 June 2014.

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. (2016). Supporting primary school teachers' classroom assessment in mathematics education: Effects on students' learning. Manuscript submitted for publication.

Wiliam, D. (2011a). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 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). 小学课堂教学形成性评价问题与对策研究 *[The problems and solutions of formative assessment of classroom teaching of primary school]* (Master's dissertation). 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: Springer.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2016a). *Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers*. Manuscript submitted for publication.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2016b). *Assessment practice and beliefs of primary mathematics teachers in China: Findings from a large-scale questionnaire survey*. Manuscript submitted for publication.

Zhong, Q. (2012). 课堂评价的挑战 [The challenges of classroom evaluation]. 全球教育展望, *1*(10), 10-16