



Chinese Primary School Mathematics Teachers' Assessment Profiles: Findings from a Large-Scale Questionnaire Survey

Xiaoyan Zhao¹  •
Marja Van den Heuvel-Panhuizen^{1,2} •
Michiel Veldhuis²

Received: 20 April 2017 / Accepted: 15 July 2017 / Published online: 1 August 2017
© The Author(s) 2017. This article is an open access publication

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

✉ Xiaoyan Zhao
x.zhao@uu.nl

¹ Freudenthal Institute, Faculty of Science, Utrecht University, Princetonplein 5, Buys Ballot Building, 3584 CC Utrecht, the Netherlands

² Freudenthal Group, Faculty of Social and Behavioural Sciences, Utrecht University, Heidelberglaan 1, Langeveld Building, 3584 CS Utrecht, the Netherlands

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

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.

Literature Review

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.

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 problem-solving 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.

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 et al., 2017). 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.

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?*

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.

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 (cf. 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."

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.

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.

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 (cf. 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 (cf. 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 (cf. 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).

Results

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 of

them (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.

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 (χ^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).

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

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

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%).

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

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

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

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%).

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.

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.

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.

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

Table 3 Factor loadings of the items on Assessment methods ($\alpha = 0.677$)

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

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

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

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 ($.092 < r < .400$).

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, Fig. 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.

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

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

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 ($\alpha = 0.691$)

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

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 Fig. 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.

Table 7 Factor loadings of the items on Perceived usefulness of assessment ($\alpha = 0.794$)

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

Table 8 Factor loadings of the items on Acceptance of assessment ($\alpha = 0.701$)

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

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

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	—			
IAES, Importance extra-curricular skills	.235	.233	.320	.098	.700	—		
PUA, Perceived usefulness 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 $p < .01$, except for (n.s.)

n.s. not significant

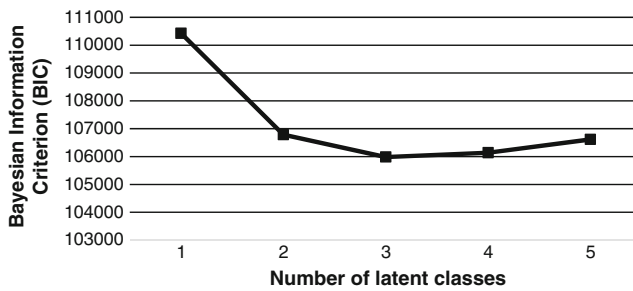


Fig. 1 The value of the Bayesian Information Criterion (BIC) for one to five latent classes

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 ($\chi^2(2, N = 1061) = 194.9, p < .001$).

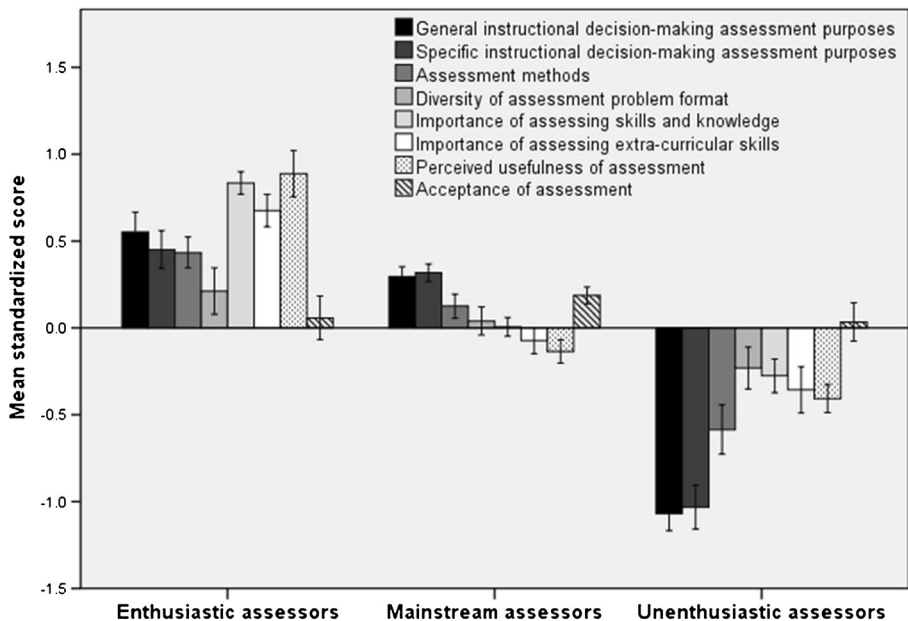


Fig. 2 Mean standardized scores on factors for teachers in the three latent classes. Whiskers indicate 95% confidence interval

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

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

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). 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 NWO MaGW/PROO: Project 411-10-750]. The authors thank Prof. 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.

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.

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: Adjudging 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' self-reported 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 biaoizhun (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 biaoizhun (2011 nian ban)* [Mathematics curriculum standards of nine-year compulsory education (2011 version)]. Retrieved from http://ncet.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 jiechuan “tongkao tongce” de zaisikao [Rethinking of standardized tests in primary education]. *Jiaoyu celiang yu pingjia (Liluban)*, 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. Oosterle, 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.