



**HAL**  
open science

# Mathematics Assessment Practices of Primary School Teachers in France

Nathalie Sayac, Michiel Veldhuis

► **To cite this version:**

Nathalie Sayac, Michiel Veldhuis. Mathematics Assessment Practices of Primary School Teachers in France. *International Journal of Science and Mathematics Education*, 2021, 20, pp.1595-1610. 10.1007/s10763-021-10229-3 . hal-04369391

**HAL Id: hal-04369391**

**<https://hal.science/hal-04369391>**

Submitted on 4 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Mathematics Assessment Practices of Primary School Teachers in France

Nathalie Sayac<sup>1</sup> · Michiel Veldhuis<sup>2,3</sup> 

Received: 2 July 2021 / Accepted: 2 October 2021  
© The Author(s) 2021

## Abstract

We investigated French primary school teachers' assessment practice in mathematics. Using an online questionnaire on teachers' background, teaching, and grading practice, we were able to determine assessment profiles of 604 primary school teachers. As evidenced by the teachers' scores on the latent factors Assessment purposes, Assessment practices, and Differentiation, teachers with the profile of *Enthusiastic assessors* view assessment as more useful and use it more often to adapt their instruction than teachers with the profile of *Unenthusiastic assessors*. This can be useful for practice and sheds more light on French teachers' assessment practices in mathematics. It is also interesting to compare the results of this survey with those from China and the Netherlands, as the differences reflect different assessment cultures and may shed light on some of the results of international large-scale assessments such as PISA.

**Keywords** Assessment · France · Mathematics education · Profiles · Primary school

Many researchers, practitioners, and policymakers agree that assessment and its practices play an important role in school systems and student achievement. In many countries, recently, reforms have taken place with the aim to promote assessment practices which assist students' study success. Every country has their own specific assessment culture, depending on the institutional, political, and historical contexts (Remesal, 2007; Schoenfeld, 2007). In France, the historical context of Jesuit colleges in which competition and high performance were central did not aid the spread of assessment practices with a formative purpose (Maulini, 2003; Merle,

---

✉ Michiel Veldhuis  
m.veldhuis@ipabo.nl

<sup>1</sup> Laboratoire de Didactique André Revuz, Université Rouen Normandie, Mont-Saint-Aignan, France

<sup>2</sup> iPabo University of Applied Sciences, Amsterdam, the Netherlands

<sup>3</sup> Freudenthal Institute, Utrecht University, Utrecht, the Netherlands

2018), but elicited a more normative vision of assessment focusing more on classifying students according to their summative assessment results than on helping them further in their learning. The subject of mathematics is used as a selection criterion in the French school system; therefore, its assessment naturally follows the normative vision on assessment, while it has repeatedly been established that using a formative approach in assessment can contribute to students' learning success (Assessment Reform Group, 1999; Black & Wiliam, 1998; Brookhart, 2003; Shepard, 2009; Veldhuis & Heuvel-Panhuizen, 2020). This raises the question of how primary school teachers actually assess their students within their mathematics classes in France? Building on the work of Veldhuis and Heuvel-Panhuizen (2014) in the Netherlands and in China (Zhao et al., 2018), we used a large-scale questionnaire to explore French teachers' assessment practices in their primary school mathematics classrooms. In this article, we present the assessment context in France, our approach to study teachers' practices, and the methodology used to analyze the results allowing us to understand teachers' assessment practices in primary school mathematics education. We will conclude by comparing our results with the findings of the earlier studies in the Netherlands and China.

## Assessment Context in France

Recently, student assessment has become an important issue for all educators until the highest levels of government, in France. This particular interest has been visible in several ways. First of all, in the Law of July 2013 (*Loi d'orientation et de programmation pour la refondation de l'École de la République*), in which new institutional rules for assessment were elaborated, focusing on "positive assessment" that valorizes progress and distances itself from sanction grading. Furthermore, the simplicity and readability for parents of the results of positive assessment of knowledge, competencies, and students' progress is underlined. By the creation of CNESCO (*Conseil national d'évaluation du système scolaire*) in 2014 and the publication of several prescriptions focusing on assessment issues (new programs in 2016, new reference levels for professional competencies, etc.), the institutional pressure on the assessment that French teachers do has become stronger related to their everyday classroom activities. What remains unclear, however, is the ways they use to really assess their students' learning.

A report published in July 2013 by the Inspectorate of education (*l'Inspection générale de l'éducation nationale (IGEN)*) described that:

Objectivity is mostly missing. No-one knows what is assessed, the levels of performance are not defined. [...] Differentiation between formative and summative assessment is missing. The mix of these two makes that teachers do not know their students' performance levels and gives students the feeling that they are continually being assessed, which is a stressful situation for some. (IGEN, 2013, p. 31).

The TALIS questionnaire (DEPP, 2014) also showed some characteristics of the French assessment situation. Notably, it was shown that over 90% of the French teachers had very diverse assessment practices and that 86% of them devised their own assessments (whereas the international average was about 68%). Concerning assessment in mathematics, even less is known, even though national assessments (CEDRE) and international assessments (TIMSS, PISA) show that for some years French students have a rather mediocre level and that there is a strong correlation between social inequity and educational inequity (Rocher, 2015).

## Research on Teachers' Assessment Practices in Mathematics

Previous research allowed us to distill some characteristics of primary school mathematics teachers' assessment practices in France (Sayac, 2016, 2017, 2019). First of all, the assessment practices of these teachers are mostly traditional and normative. The teachers mainly assess their students summatively, with some variations in timing and frequency (after each teaching sequence, every period, intermediary). They mostly conceive their assessments alone, using a variety of resources, such as handbooks or websites, that they combine to design tests that suit their classes.

The mathematical tasks that are included in these assessments have a rather low level of complexity (they are easy), so as to not let students fail. The tasks correspond almost exclusively to the tasks that students had already solved during the lessons before the assessment. The teachers often wait to be sure that the majority of the students understand the lesson before proceeding with its assessment. Upon analyzing the mathematical tasks used for assessment purposes with a didactical tool that was designed to take into account different elements of complexity (Sayac & Grapin, 2015), it was found that French teachers make personal choices concerning the assessment tasks by referring to their vision on mathematics (often very segmented) and their teaching. This is related to the work of Cizek et al. (1996) who showed that teachers' assessment practices were often very variable and unpredictable, because they depend on many personal and contextual parameters.

## Didactical Context for Assessment in Mathematics

To study assessment in mathematics and its ordinary practices, Sayac (2017) developed a didactical framework that allows to study different *assessment episodes* teachers use in their classes and their accompanying *assessment logic*. The assessment episode has been devised to leave the traditional functions of assessment, formative and summative, behind and explicitly take into account the disciplinary content in its assessment. To study an assessment episode, one has to focus on its timing, the nature of its mathematical tasks, its management (in terms of judgments and regulation), but also the didactical contract (Brousseau, 2006). The *assessment logic* of the teacher that proposes assessment episodes to the students concerns a number of indicators that can describe this personal

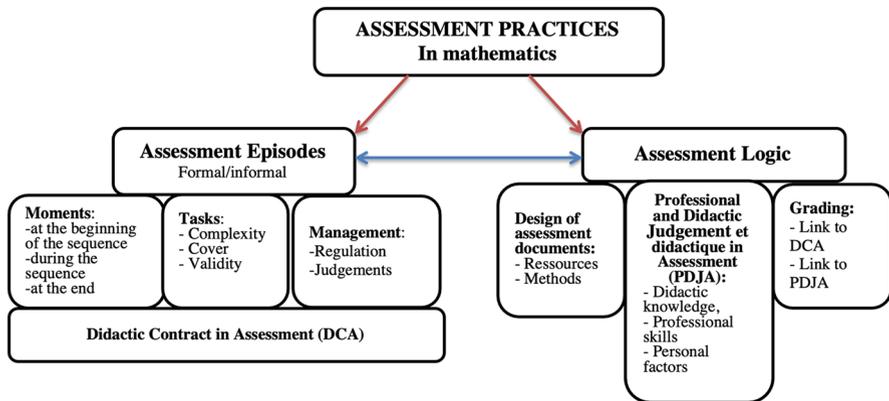


Fig. 1 Elements of teachers' assessment practice

logic, notably by the way teachers design their assessment documents, their particular mathematical knowledge or skills, their professional and didactical judgment, and the grading.

This *professional and didactical judgment in assessment* is inspired by work in educational sciences on professional assessment judgment (Laveault, 2008; Mottier Lopez & Allal, 2008), but is specific for mathematics as it integrates notions that were developed within the field of didactics of mathematics, such as didactical vigilance (Pézard, 2010). This judgment depends on mathematical and didactical knowledge, on teachers' knowledge on assessment—their assessment literacy (Webb, 2002; White, 2009)—and on individual factors such as beliefs and ideas on learning, on assessment, and experiences with assessment.

The scheme in Fig. 1 illustrates the elements that were taken into account to study teachers' assessment practices.

To better understand what is going on in the assessment of the mathematics learning of French primary school students, we focused on the teachers' assessment logic. The quantitative study we carried out aimed to find how the different indicators for the teachers' assessment logic as it has been defined above are related.

Main research question: what does the assessment practice of French primary school teachers in mathematics look like?

To answer this research question, we explored the elements that constitute teachers' professional judgment and assessment didactics by relating their personal characteristics (education, ideas about mathematics and assessment) to their assessment and grading practices (assessment methods, grading, sources for designing assessments).

This information can be useful to find out how to assist pre- and in-service teachers in using assessment in a more professional way when they assess their students' mathematics learning.

## Method

### Participants

Six hundred four primary public-school teachers from the same academic district<sup>1</sup> (académie de Créteil) completed an online questionnaire. This questionnaire was sent out via professional lists, usually used to disseminate institutional information. Teachers were not compensated in any way for filling in the questionnaire. This number is not representative<sup>2</sup> of all primary school teachers in the district, but it is still high enough to inform us about assessment practices of these teachers.

### The Questionnaire

With this perspective, we devised a questionnaire consisting of three parts, allowing us to gather the necessary information to answer our questions about French primary school teachers' assessment practices and beliefs.

#### Part I: Who Are You? (13 Questions)

Questions related to the teachers' background (gender, age, professional experience, grade level) and questions about teachers' perspective and feelings on mathematics. The latter information is used to determine teachers' professional and didactical judgment about assessment, depending on their mathematical and didactical knowledge.

#### Part II: Assessment Practice (13 Questions)

Questions focusing on identifying the teachers' assessment logic, by asking about the role of assessment in their practice, their ideas about positive assessment, particularly their professional judgment and didactics related to assessment. Also, the frequency of their mathematics assessment, the type of assessment (summative, formative, diagnostic, self- or peer-assessment), and the ways they differentiate were asked.

#### Part III: Students and Grading (12 Questions)

Questions about teachers' grading practices towards students, parents, and the institution. Finally, we asked teachers what kind of difficulties they encounter when assessing their students' mathematics learning.

---

<sup>1</sup> In France, there are 26 metropolitan academic districts and 4 overseas academic districts.

<sup>2</sup> In 2019, the district of Creteil had 29,880 primary school teachers teaching in public schools.

## Data Analysis

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 characterize these French teachers' assessment practice and determine their assessment profiles (this same approach was used in Veldhuis & Heuvel-Panhuizen, 2014; Zhao et al., 2018).

The input for the factor analysis was teachers' answers to the 25 questions focusing on how teachers view their assessment and grading practice. We used two latent variable modeling techniques to identify the latent structure of the items and teachers' responses to these in the questionnaire. To decide about the most appropriate models, we used substantive and 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 performed 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.

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

Finally, differences between teachers with the different assessment profiles on a number of background variables were investigated with analyses of variance (ANOVA), Kruskal–Wallis, and  $\chi^2$ -differences tests. With these analyses, the defining elements for each profile could be determined. The inferential analyses were performed in SPSS 24 and all latent variable modeling in MPlus 6 (Muthén & Muthén, 1998–2010).

## Results

### Factor Analysis

After comparing one- to ten-factor solutions, our exploratory factor analyses delivered a four-factor solution that had a good enough fit ( $\chi^2$  (1949,

**Table.1** Factor loadings of the items on Assessment purposes ( $\alpha=0.459$ )

Questionnaire item	Factor loading
Role assessment: determine students' level	0.655
Positive assessment is adapted to students' level	0.522
Role assessment: institutional	0.516
Role assessment: help in planning teaching	0.431
Mark students: 0–20	0.406
Construct assessment using class guide	0.424
Construct assessment using teacher's own trajectory	0.409
Construct assessment using teacher guide	0.407
Construct assessment using internet sources	0.351
Positive assessment deals with students' well-being	0.326
Construct assessment using competencies in the curriculum	0.322
Role assessment: fill in students' portfolio	0.317
Role assessment: determine students' progress	0.312
Assess students by observing them during entire teaching sequence	0.306

**Table.2** Factor loadings of the items on Assessment practice ( $\alpha=0.324$ )

Questionnaire item	Factor loading
Assess students when a teaching sequence is finished	0.590
Role assessment: fill in students' portfolio	0.579
Use summative assessment	0.567
Use self-assessment	0.503
Use differentiated assessments	0.498
Use formative assessment	0.487
Role assessment: adjust teaching to students' needs	0.484
Positive assessment promotes students' self-image	0.454
Role assessment: plan instruction	0.433
Construct assessment using class guide	0.430
Assess students at another moment	0.394
Overall level of students in mathematics	0.314
Failures are included in students' portfolios	0.303

$N = 599$ ) = 3899.1,  $p < 0.0001$ ,  $RMSEA = 0.041$ ,  $CFI = 0.78$ ). Also, these four factors all had eigenvalues over 3. The  $\chi^2$ -statistic of the overall model fit was significant, which indicates a model with a less than optimal fit. Nevertheless, this nested four-factor solution fitted significantly better than the three-factor solution, as illustrated by the Satorra-Bentler scaled  $\chi^2$ -test (TRD ( $df = 64$ ) = 361.4,  $p < 0.0001$ ). Most of the items in the questionnaire that could be expected to form a scale, due to their content and focus, loaded coherently on different latent factors, thus providing substantive evidence for this four-factor solution (see Tables 1, 2, 3, and 4 for the items constituting the latent factors).

**Table.3** Factor loadings of the items on Differentiation ( $\alpha=0.514$ )

Questionnaire item	Factor loading
Differentiation: by providing less complex tasks	0.835
Differentiation: by providing individualized help (instruments)	0.564
Differentiation: by providing fewer tasks	0.558
Differentiation: by providing individualized help (teacher)	0.532
Differentiation: by providing help in reading the tasks	0.526
Differentiation: by providing other help	0.505

**Table.4** Factor loadings of the items on Grading practice ( $\alpha=0.237$ )

Questionnaire item	Factor loading
Mark students: 0–10	0.798
Mark students: 0–20	0.590
Mark students: rubric (accomplished, needs work, not accomplished)	0.540
Adapt scoring procedure to the class' level	0.359

**Table.5** Correlations between factors

	Factor 2: Assessment practice	Factor 3: Differentiation	Factor 4: Grading practice
Factor 1: Assessment purposes	0.385	0.265	0.075
Factor 2: Assessment practices	–	0.323	0.054
Factor 3: Differentiation		–	0.084

All Cronbach's alphas are rather low, indicating suboptimal scales; however, the items in the scales fit very well substantially and based on the results of the factor analysis. Regarding the items that constitute these factors, we therefore decided on the following names. Factor 1, which included items the purposes or roles teachers attribute to assessment and different ways of constructing assessments, was called "Assessment purposes." Factor 2, which included items on the assessment methods and types of assessments teachers use was called "Assessment practice." Factor 3, which contained only items related to teachers' differentiation practices, was called "Differentiation." Finally, factor 4, which contained items on teachers' grading practices, was called "Grading practice." Correlations between the different factors are displayed in Table 5.

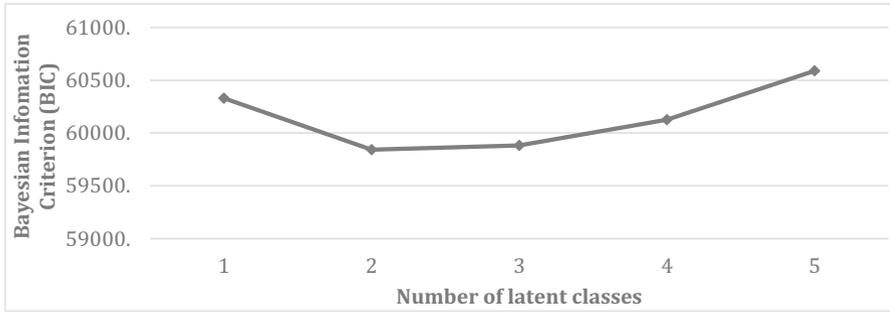


Fig. 2 BIC and number of latent classes

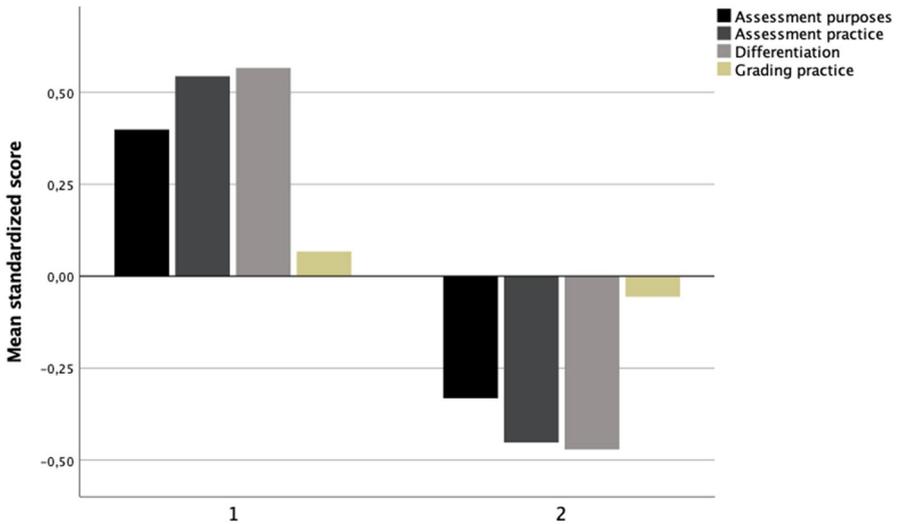


Fig. 3 Assessment profiles

### Latent Class Analysis

We used the Bayesian information criterion (BIC) and the relative entropy to select the number of latent classes the best summarizes the variation in the data. As can be seen in Fig. 2, the BIC decreased until two latent classes and then increased again. This is an indication for the two-factor solution having the better fit, because a lower BIC indicates a better fit. Additionally, the relative entropy of 0.82 of the two-factor solution was rather high.

To investigate teachers’ assessment profiles, we combined the results of the factor analysis, with the four factors of items, and the latent class analysis, with two classes of teachers. This resulted in Fig. 3 showing the different means of

teachers with the two assessment profiles on the four factors Assessment purposes, Assessment practice, Differentiation, and Grading practice.

In Table 6, we describe the differences between the teachers with the different assessment profiles on a number of relevant indicators.

Teachers with the two assessment profiles can be distinguished by their scores on three factors that are related to their assessment logic: Assessment purposes, Assessment practice, and Differentiation. The teachers in the first group, which we called the *Enthusiastic assessors*, have a vision of assessment in which it is more useful, than teachers in the second group, the *Unenthusiastic assessors*. In the view of the *Enthusiastic assessors*, assessment is more aimed at establishing students' progress (78.3% vs 53.8%), determining the content of the lessons (46.3% vs 19.3%) and adapting further instruction (89% vs 60.6%). They are also more prone to develop their assessments by themselves (62.5% vs 35.2%) or on the basis of competences listed in the study program (81.6% vs 66.1%). Concerning differentiation, the *Enthusiastic assessors* use it more often (always: 12.5% vs 3.4%, often: 39.7% vs 11.3%, never: 0% vs 27.5%).

If we look more closely to the elements that contribute to the differences between the three factors, it is interesting to see that the first difference is that of the experience teachers have with teaching (Q3). The *Unenthusiastic assessors* contain notably more teachers with relatively little experience (less than 3 years). This characteristic is emblematic for the fact that primary school teachers' pre-service education contains little to no preparation for the use of assessments. Starting teachers have to use their own previous experiences as students to shape their professional practice, which is far from ideal. They reproduce assessment practices they were used to as students, which are often rather weak in terms of usefulness for learning. The *Unenthusiastic assessors* also use less often formative assessment (Q18), peer assessment (Q20), and self-assessment (Q21). Following this same logic, they also assess less frequently their students' learning as they are very few in doing this daily or weekly (Q16). Another clearly distinctive characteristic is the way the teachers in the two groups view mathematics. The *Enthusiastic assessors* are more often "à l'aise" (at ease, find it easy) with mathematics than the *Unenthusiastic assessors* (39% vs 29.7%).

There are also a number of characteristics on which these two groups of teachers do not differ, but that are rather interesting because they can be considered to be representative for the assessment practice of teachers in France. Among these characteristics it clearly comes forward, in line with earlier studies (Sayac, 2016), that summative assessment is deemed very important (50% of the teachers in both groups "always"), that they have a rather individual conceptualization of assessment and a way of assigning grades that indicates the level of competence of students instead of letters or numbers (about 75% of teachers in both groups). The teachers also were in agreement when we look at how they gauge the level of their students' mathematical competencies (Q12: level "weak" for 46% of the teachers and "satisfactory" for 43%). Neither age, gender, nor previous studies were related to the assignment to either group. This means that the teachers' assessment logic is more or less independent of these characteristics and that it reflects

**Table 6** The means and differences for the two assessment profiles on relevant variables

	Total	Assessment profiles		Significant differences
		Group 1 Enthusiastic	Group 2 Non-enthusiastic	
Assessment purposes	–	0.399 (1.014)	–0.332 (0.858)	1 > 2
Assessment practice	–	0.544 (0.847)	0.453 (0.886)	1 > 2
Differentiation	–	0.566 (0.756)	–0.471 (0.932)	1 > 2
Grading practices	–	0.067 (1.056)	–0.056 (0.949)	n.s
<i>N</i>	599	272	327	n.s
Age (mode)	31–40 years	41–50 years	31–40 years	n.s
Gender (% female)	80.8	82	79.8	n.s
Teaching experience (mode)	10–20 years	10–20 years	<3 years	n.s
Professional development (%yes)	9	14	4	1 > 2
Stance towards mathematics (mean)	3.03	3.16	2.92	n.s
Difficult (%)	10.7	7.7	13.1	n.s
Indifferent (%)	9.7	7.7	11.3	n.s
Not problematic (%)	45.7	45.6	45.9	n.s
Easy (%)	33.9	39	29.7	1 > 2
Frequency assessment				
Daily	11.9	20.9	4.9	1 > 2
Weekly	14.3	20.1	9.8	1 > 2
Two weekly	28.9	24.3	32.6	n.s
Monthly	28.2	23.4	31.9	n.s
Periodically	13.2	9.6	16	n.s
Trimestrally	3.5	1.7	4.9	n.s
Formative assessment (mode)	Sometimes	Always	Sometimes	n.s
Never	8.9	3	13.9	1 < 2
Sometimes	11.5	3.7	17.9	1 < 2
Often	29.7	26	32.7	n.s
Always	27.3	42.8	14.5	1 > 2
Summative assessment (mode)	Always	Always	Always	n.s
Always	49.5	48.9	50	n.s
Peer assessment (mode)	Never	Often/sometimes	Never	n.s
Never	46.7	22.4	67.2	1 < 2
Sometimes	21	26.5	16.4	n.s
Often	23.1	34.3	13.6	1 > 2
Self-assessment (mode)	Never	Often/sometimes	Never	n.s
Never	29.4	9.6	46.1	n.s
Sometimes	29.1	36.5	22.7	1 > 2
Often	12.3	19.6	6.2	n.s
Always	3.9	7.4	0.9	n.s

Note: In bold indicates the significantly higher value

an individual style of teaching or assessing that does not reflect the teacher's age, gender, or previous studies.

## Summary of Findings

Teachers' previous studies have less influence on their assessment practices than their views on mathematics. The teachers that are more "à l'aise" with mathematics are more often *Enthusiastic assessors* than teachers that have difficulty with mathematics (whom are more often *Unenthusiastic assessors*). These *Enthusiastic assessors* are also more open towards different forms of assessment, such as formative, peer, and self-assessment. Nonetheless teachers in the *Unenthusiastic assessors* report to not have more trouble with the assessment of a variety of mathematics domains than the *Enthusiastic assessors*.

The beliefs and representations of teachers about assessment and about learning affect their assessment practices and the way they differentiate between students. *Enthusiastic assessors* have a more positive view of assessment that shares elements with assessment for learning (Broadfoot et al., 2002; Klenowski, 2009) than *Unenthusiastic assessors*. As such, *Enthusiastic assessors* consider positive assessment to be an assessment that aims at helping students in their learning (76.5% vs 50.8%), that takes students' well-being into account (23.9% vs 8%) and focuses on students' self-worth (61% vs 36.7%). This appears to be related to how the teachers differentiate. The *Enthusiastic assessors* differentiate much more often in their assessments (52% vs 15%) than the *Unenthusiastic assessors*. How the teachers differentiate also appears to be related to their assessment profile: with *Enthusiastic assessors* choosing more often to differentiate their assessments in terms of alleviating the complexity of the proposed questions (58.5% vs 32.1%), by reducing the number of questions (60.3% vs 37%), by helping with reading the questions (80.9% vs 47.7%), or by proposing supportive tools to their students (73.9% vs 28.7%) or individual help (51.1% vs 26.6%).

To design their assessments, *Enthusiastic assessors* refer more often to the competences as described in the lesson programs (81.6% vs 66.1%). Their more positive attitude about their own mathematical knowledge probably leads them to more often use their own ideas about learning progressions to design their assessments (62.5% vs 35.2%). Using internet sources for their assessments is equally often done.

## Discussion

The results of our statistical analyses into the responses to our questionnaire allow us to better understand French primary school teachers' assessment practices in their mathematics classes. We distinguished two assessment profiles, with each containing about half of the teachers: 45.5% in the first and 54.6% in the second. By considering the relevant indicators that can characterize these groups, we qualified the first group as the *Enthusiastic assessors* and the second as the *Unenthusiastic assessors*. This characterization allows us to draw parallels with the earlier studies in the

Netherlands and China (Veldhuis & Heuvel-Panhuizen, 2014; Zhao et al., 2018) on how primary school mathematics teachers use and perceive assessment. Effectively, although the questions teachers answered were not exactly the same and the Dutch, French, and Chinese contexts are quite different, these three studies strived for the same goal of the identification of teachers' assessment profiles on the basis of questionnaire data on teachers' reported assessment practice with the same analytical approach. Several of the indicators are exactly the same such as the age, gender, educational background, and teaching experience of the primary school teachers and the list of possible assessment purposes and methods. In the three studies, 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.

In the Dutch study, analyzing about 1000 teachers' responses, four assessment profiles were distinguished (*Enthusiastic assessors* 28.5%, *Mainstream assessors* 35.3%, *Unenthusiastic assessors* 25.8%, and *Alternative assessors* 10.3%), while in the Chinese study three profiles were found (*Enthusiastic assessors* 21.7%, *Mainstream assessors* 53.1%, and *Unenthusiastic assessors* 25.2%), and finally in the French study two groups of teachers could be identified (*Enthusiastic assessors* 45.4% and *Unenthusiastic assessors* 54.6%). The different number of profiles found in the studies is probably due to the exact nature of the questions in the different questionnaires and other cultural, educational, and contextual parameters which are difficult to pinpoint. Nonetheless several of the findings appear to be related to the overlapping questions in the different questionnaires.

In the three countries, teachers who were qualified as *Enthusiastic assessors* have a more or less positive stance towards assessment and use it more often in their practice. Interestingly the relation of their assessment profile with age and professional experience differs between the countries. In China there were no significant differences on these indicators between teachers with the different assessment profiles, whereas the *Enthusiastic assessors* in France were significantly older and had more experience and in the Netherlands were significantly younger and had less experience. This surprising difference between the countries could be related to the way teachers are educated in these countries. In France, the mathematics content in initial teacher training is very limited (Sayac & Grapin, 2015) and has very little focus on assessment, which leads to teachers who, at the start of their career, have a rather negative stance towards assessment and do not see it as a useful tool to assist their students' learning. In the Netherlands, the education for teachers in primary school takes 4 years. Similar to France, assessment, formative or positive, is not a substantial factor in the teacher education curriculum.

## Limitations

Evidently our study has some limitations. The number of respondents and the representativity of our sample are a concern. Even though over 600 teachers is an important number, they were all from the same district and were thus not representative

for teachers in French primary education. Social desirability of the answers to the questions might have had an influence too, specifically linked to formative assessment as Issaieva et al. (2011) found. Also, we have found that assessment practices are often close to the heart of teachers, leading to them not always responding truthfully to questions concerning their own classroom practice (Sayac, 2017). Studying professional practices with a questionnaire allows to obtain a general view of these practices but does not lead to really understanding what happens within the classrooms and in teachers' minds (the so-called black box, Black & Wiliam, 1998). Not all the dimensions of the assessment logic of primary school teachers have been taken into account in this study, notably how teachers articulate assessment episodes and those that are not and what *assessment arrangements* they exactly employ in their classrooms when grading or teaching their students (Merle, 2007). These dimensions are essential to determine how assessment is useful for both teachers and students. Other aspects of professional judgment and didactics in assessment like mathematical knowledge for teaching (Ball et al., 2005, 2008; Clivaz, 2012) and teachers' stance towards mathematics to know how these factors really affect teachers' assessment practice in their classrooms. To investigate these questions, a further observational study would be in order with a more qualitative analysis.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. University of Cambridge School of Education.
- Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 1422, 43–46.
- Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42, 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Hareln, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. University of Cambridge School of Education.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Brousseau, G. (2006). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990 (Vol. 19)*. Springer Science & Business Media.

- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159–179.
- Clivaz, S. (2012). Connaissances mathématiques de l'enseignant et bifurcations didactiques: Analyse d'un épisode [Mathematical knowledge of the teacher and didactic bifurcations: Analysis of an episode]. *Recherches En Didactiques*, 2, 29–45.
- Conseil national d'évaluation du système scolaire (CNESCO). (2014). *L'évaluation des élèves par les enseignants dans la classe et les établissements: Réglementation et pratiques. Une comparaison internationale dans les pays de l'OCDE* [Student assessment by teachers in the classroom and schools: Regulations and practices. An international comparison in OECD countries]. Author.
- 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). Springer.
- Direction de l'évaluation, de la prospective et de la performance n°23 (DEPP). (2014). *Enseignant en France: Un métier solitaire*, France, Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche [Teacher in France: Un profession solitaire, France, Ministry of National Education, Higher Education and Research].
- 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. <https://doi.org/10.1080/10705519909540118>
- Inspection Générale de l'Éducation Nationale (IGEN). (2013). *La notation et l'évaluation des élèves éclairées par des comparaisons internationales, Rapport n°2013- 072* [Pupil Scoring and Assessment Informed by International Comparisons, Report No. 2013-072. France, Ministry of National Education]. France, Ministère de l'Éducation nationale.
- Issaieva, É., Pini, G., & Crahay, M. (2011). Positionnements des enseignants et des élèves du primaire face à l'évaluation: Une convergence existe-t-elle? [Positioning of teachers and primary school students with regard to assessment: Does convergence exist?]. *Revue française de pédagogie. Recherches En Éducation*, 176, 5–26.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 263–268.
- Laveault, D. (2008). Le jugement professionnel: Foyer de tensions et de synergies nouvelles en évaluation scolaire [Professional judgment: Hotbed of new tensions and synergies in academic evaluation]. *Revue Suisse Des Sciences De L'éducation*, 30(3), 483–500.
- Maulini, O. (2003). L'école de la mesure. Rangs, notes et classements dans l'histoire de l'enseignement [The school of measurement. Ranks, marks and rankings in the history of education]. *Educateur, n°mars 2003. Un Siècle D'enseignement En Suisse Romande*, 2, 33–37.
- Merle, P. (2007). *Les notes* [The notes]. Presses Universitaires de France, Paris.
- Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives* [School assessment practices. History, difficulties, prospects]. Presses Universitaires de France, Paris.
- Mottier Lopez, L., & Allal, L. (2008). Le jugement professionnel en évaluation: Un acte cognitif et une pratique sociale située [Professional judgment in evaluation: A cognitive act and a situated social practice]. *Revue Suisse Des Sciences De L'éducation*, 3, 465–482.
- Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, 8(3), 369–377. <https://doi.org/10.1037/1082-989X.8.3.369>
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide. Sixth edition*. Muthén & Muthén.
- Pézar, M. (2010). Installer la paix scolaire, exercer une vigilance didactique [Establish school peace, exercise didactic vigilance]. *Recherches En Didactique Des Mathématiques*, 30(2), 197–261.
- Remesal, A. (2007). Educational reform and elementary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *The Curriculum Journal*, 18(1), 27–38.
- Rocher, T. (2015). Mesure des compétences – Méthodes psychométriques utilisées dans le cadre des évaluations des élèves [Competency measurement - Psychometric methods used in student assessments]. *Education & Formation, No. 86–87*, 37–61.
- Sayac, N. (2017). Étude des pratiques évaluatives en mathématiques des professeurs des écoles en France: Une approche didactique à partir de l'analyse des tâches données en évaluation [Study of evaluative practices in mathematics of school teachers in France: A didactic approach based on the analysis of the tasks given in evaluation]. *Revue Mesure Et Évaluation En Éducation*, 40(2), 1–31.

- Sayac, N., & Grapin, N. (2015). Évaluation externe et didactique des mathématiques: Un regard croisé [External and didactic assessment of mathematics: A crossed look]. *Recherches En Didactique Des Mathématiques*, 35(1), 101–126.
- Sayac, N. (2016). *How are pupils in French primary school assessed in mathematics? A didactical approach to explore this question*. ICMI Study 13. Hamburg (Germany).
- Sayac, N. (2019). Approche didactique de l'évaluation et de ses pratiques en mathématiques [Didactic approach to assessment and its practices in mathematics.]. *Recherches En Didactique Des Mathématiques*, 39(3), 283–329. <https://revue-rdm.com/2019/approche-didactique-de-levaluation-et-de-ses-pratiques-en-mathematiques/>
- Schoenfeld, A. H. (2007). *Assessing mathematical proficiency* (Vol. 53). Cambridge university press.
- Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, 28(3), 32–37.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014). Primary school teachers' assessment profiles in mathematics education. *PLoS One*, 9(1), e86817.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2020). Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement. *Mathematics Education Research Journal*, 32(3), 449–471. <https://doi.org/10.1007/s13394-019-00270-5>
- Webb, N. (2002, April). *Assessment literacy in a standards-based urban education setting*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana.
- White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course. *Asian EFL Journal*, 33(1), 1–36.
- Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2018). *International Journal of Science and Mathematics Education*, 16(7), 1387–1407.