Author(s): Nieminen, Juuso Henrik; Atjonen, Päivi; Tolvanen, Asko

Title: Diversity of assessment practices and student perceptions of assessment in Finland

Year: 2024

Version: Published version

Please cite the original version:

Check for updates

ORIGINAL ARTICLE

The Curriculum Journal ◼BERA

Diversity of assessment practices and student perceptions of assessment in Finland

Juuso Henrik Nieminen[1] 🆔 🐦 | **Päivi Atjonen**[2] 🆔 | **Asko Tolvanen**[3] 🆔

[1]Faculty of Education, The University of Hong Kong, Hong Kong, Hong Kong

[2]The Philosophical Faculty, University of Eastern Finland, Joensuu, Finland

[3]Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

Correspondence
Juuso Henrik Nieminen, The University of Hong Kong, Pok Fu Lam, Hong Kong.
Email: juuso@hku.hk

Abstract

Educational policies and curriculum documents largely emphasise the idea that assessment should be *diverse*: that teachers should use versatile assessment practices such as tests, self- and peer-assessment and portfolios instead of mainly drawing on uniform practices, such as examinations. In this study, we examined the diversity of assessment in the low-stakes assessment culture of Finland where the national curriculum emphasises the ethos of Assessment for Learning. We were interested in how the diversity of assessment (or the lack of thereof) is related to student perceptions of assessment. We utilised both multilevel and person-oriented methods in our analysis of a national evaluation dataset of 187 Finnish primary and lower secondary teachers and their 2370 students. First, we noted that according to the teachers' responses, classroom assessment in Finland is not particularly diverse. At the same time, students did not find assessment particularly helpful for their learning. Second, a multilevel analysis revealed varying correlations between the prevalence of various assessment practices and students' perceived usefulness of the corresponding practices. Third, we explored teacher subgroups in terms of their 'assessment menu' through a latent class analysis. All identified teacher subgroups drew heavily on closed-book examinations. Finally, we analysed whether the assessment perceptions of students differ according to their teacher's latent class. The students whose teachers used the most diverse assessment menu reported the most

pressure in assessment. We suggest these findings stem from the socio-historical context of pedagogical conservatism in Finland, combined with the low stakes of assessment. We discuss the importance of widening teachers' assessment practices in order to promote sustainable student perceptions of assessment, yet such assessment diversity must be valued in the educational system, and it needs to be based on sound pedagogical design.

INTRODUCTION

It is widely agreed in scholarly research and education policy that assessment should be diverse and versatile (e.g., O'Neill & Padden, 2022; Shepard, 2001; Warwick et al., 2015). For example, the Finnish National Core Curriculum (FNAE, 2020) mandates that students have the right to be assessed through diverse practices such as self- and peer-assessment, tests, portfolios and group projects (p. 5). Moreover, it is stated that a single assessment practice cannot meaningfully assess all learning objectives (FNAE, 2020, p. 5). Similarly, scholarly literature has suggested that instead of focusing on individual assessment practices, there is a need to examine the 'assessment menus' of teachers—the diversity of their assessment practices as a whole (e.g., Iannone & Simpson, 2022). Shulman (1987) used this metaphor to denote how diverse assessment menus should consist of nutritious ingredients: 'We are, as it were, stocking the board's pantry with a variety of completed dishes and separate ingredients from which its members will be able to fashion an assessment menu' (p. 38). At present, the ability to design and implement versatile assessment practices is seen to be at the core of teacher assessment literacy (DeLuca & Lam, 2014; Liu et al., 2023; Siegel & Wissehr, 2011; Yan & Pastore, 2022).

The key idea behind the diversity of assessment practices is that different techniques allow different forms of student learning to become visible. In education cultures centred on restricted assessment menus, students might not have the opportunity to showcase all their skills and capabilities (Black et al., 2011; Bourke & Mentis, 2014; Graham et al., 2018; Watt, 2005). The 'dominance of examinations' (Richardson, 2022, p. 22) in particular has been shown to narrow assessment menus on a global scale (see Barnes et al., 2000; Birenbaum et al., 2015; Volante et al., 2020). This issue might be particularly pertinent in high-stakes assessment cultures where the implementation of diverse assessment practices, such as self- and peer-assessment and portfolios, is reportedly met with resistance by students and teachers alike (Ratnam-Lim & Tan, 2015; Yan & Brown, 2021).

Yet, simply providing a large diversity of assessment practices without a meaningful pedagogical purpose and structure is not appropriate in itself (Chu, 2014). For example, if students perceive a diverse assessment menu as significantly increasing their workload without a clearly stated pedagogical purpose, versatile assessment might lead to instrumental learning (Gijbels & Dochy, 2006; Johansson et al., 2022). At the same time, providing a rich, diverse menu of assessment practices might foster students' reflexivity on how and why assessment is conducted (Nieminen & Lahdenperä, 2024).

In this study, we examine how the diversity of assessment practices (or the lack of thereof) is related to student perceptions of assessment in Finnish primary and lower secondary

education. Our study is conducted in Finland, which provides an intriguing context due to its low-stakes assessment culture. In Finland, there are no high-stakes testing procedures at these levels of education, and there is a strong trust in teachers' professional autonomy to design classroom assessment practices as they wish. To examine the connection between teachers' assessment practices and students' perceptions of assessment, we conduct both multilevel and person-oriented quantitative analyses based on a national evaluation dataset. While student perceptions of assessment have enjoyed vigorous research interest (see, e.g., Struyven et al., 2005; Van der Kleij & Lipnevich, 2021), studies have rarely examined teacher and student views on assessment simultaneously (Pat-El et al., 2015; Veugen et al., 2021). As Wurf and Povey (2020) observed, the learning potential of assessment depends not only on teacher perceptions and practices but also on student perceptions. Any implementation of formative assessment or Assessment for Learning practices is successful only if students perceive these practices as meaningful and purposeful (Brown, 2022). Thus, we examine students' perceptions of assessment in relation to their teachers' assessment practices.

DIVERSE ASSESSMENT FROM THE VIEWPOINTS OF TEACHERS AND STUDENTS

In this study, we approach assessment as a 'menu' (as phrased by Shulman, 1987), as opposed to investigating individual assessment designs or practices. The diversity of assessment offers a meaningful way to examine assessment systems as a whole, as any individual assessment practice—be it a closed-book examination, a self-assessment sheet or an e-portfolio—might either hinder or support student learning in a given context and situation.

The Finnish education system has received extensive international attention in traditional media and research, yet this attention has largely focused on large-scale assessment results such as Programme for International Student Assessment (PISA) measurements (Thrupp et al., 2023). At the same time, classroom assessment has been relatively understudied in Finland (Kumpulainen & Lankinen, 2016). This area has recently been explored through disciplinary approaches but not yet through broad, system-level investigations (see, e.g., Nieminen & Atjonen, 2022; Mäkipää & Ouakrim-Soivio, 2019). In this section, we provide background literature on how earlier studies have mapped out teacher assessment practices and student perceptions of these practices.

Diverse assessment as teachers' responsibility

The idea of diverse assessment has been widely promoted in education policy and practice. In their review, Fulmer et al. (2015) formulated a three-fold theory to understand the factors that influence the 'assessment menu' of teachers, shedding light on micro- (e.g., teacher assessment literacy and values), meso- (e.g., school cultures and opportunities for professional development) and macro- (e.g., wider assessment policies and cultures) systems. Our study focuses on a macro-level view of a low-stakes culture in Finland; this culture arguably surrounds student perceptions of assessment as well.

While our study focuses on Finland's low-stakes assessment culture, we outline some crucial global macro-level trends in classroom assessment. First, even though formative assessment has been advocated for in educational research, policy and practice for decades, its implementation remains stubbornly tricky in both high- and low-stakes assessment cultures due to the influence of high-stakes testing (Birenbaum et al., 2015; Stobart, 2008; Volante et al., 2020). Any investigation of teachers' assessment menus may need to acknowledge

the gold-standard status of testing in educational assessment (Richardson, 2022). In classroom assessment, examinations have often been noted as the most used practice in many contexts (e.g., Cotton, 2017; Kippers et al., 2018).

Studies on teacher perceptions of assessment have unpacked the tensions teachers must live with as they juggle the accountability and learning purposes of assessment (Chan & Tan, 2022; Harris & Brown, 2009; Veldhuis & van den Heuvel-Panhuizen, 2014; Winstone & Carless, 2021). Research literature has listed various barriers to teachers seeking to make their assessment practices more diverse, such as teachers' lack of assessment literacy (Siegel & Wissehr, 2011), the washback effect of high-stakes testing (Chan, 2020) and lack of time and resources (O'Neill & Padden, 2022).

Taking an 'assessment menu approach', we are interested in the prevalence of various assessment practices that teachers may use. For this purpose, clustering and profiling methods are suitable. While teachers and pre-service teachers have been clustered and profiled in terms of their teaching styles and practices (e.g., Kim & Son, 2021; Yang & Hu, 2019) and in terms of their assessment conceptions (e.g., Kyttälä et al., 2024; Lin et al., 2024) and approaches (DeLuca et al., 2021), we were unable to identify many studies that have used profiling techniques for data on teacher assessment *practices*, as we do in this study. Some relevant studies have been conducted in tertiary rather than primary and secondary education (Picos et al., 2013; Picos & Lopez-Pastor, 2013). For example, Paakkari et al. (2022) clustered Finnish health education teachers in terms of their assessment conceptions and practices, ending up with three clusters: problematic assessors (47.3% of the sample), learning supportive assessors (27.3%) and norm-based assessors (25.5%).

Student perceptions of the diversity of assessment

Our premise is that the assessment practices that students experience play a potential role in how students perceive assessment. Being exposed to formative assessment—and receiving adequate training on how to make the most of it—might increase positive perceptions of formative assessment (as was hinted by Nieminen & Lahdenperä 2024). On the other hand, if students are exposed mainly to a unilateral set of assessment practices, such as in test-driven assessment cultures, they might not perceive formative assessment practices as helpful (Nieminen & Atjonen, 2022). As Wurf and Povey (2020) pointed out, students may have a limited understanding of how assessment and learning are related and how students should contribute to this relationship (see also Hannigan et al., 2022).

Student perceptions and conceptions of assessment have received vigorous research interest during the two recent decades (Brown, 2022; Struyven et al., 2005; Van der Kleij & Lipnevich, 2021). Research on feedback has shed light on assessment methods as well (e.g. Strijbos et al., 2021). For example, Brooks et al. (2019) reported that Australian students saw improvement-oriented feedback as most helpful to learning. Students in a Norwegian study by Vattøy et al. (2022) revealed that dialogic interaction was crucial in using digital feedback in upper secondary school: through such interactions, students developed their understanding of assessment and thus their perceptions of it. A review by Van der Kleij and Lipnevich (2021) indicated that student perceptions of feedback varied tremendously, but notably, students' perceptions did not unambiguously accord with their teachers' intentions regarding how useful and effective feedback was. Students reportedly opined that classroom climate, peers' ability to express their comments, and the timing of feedback and its modality (written, verbal, audio, individual and group-based) were important for progress.

The need to understand student and teacher views simultaneously

As Monteiro et al. (2021) and Alkharusi (2010) have noted, teachers' assessment practices affect how students perceive assessment. However, teachers' and students' viewpoints on assessment have been scarcely studied together, particularly when it comes to large-scale investigations (as argued by Pat-El et al., 2015). We identified a few key references on this front.

Daly et al. (2012) reported from the UK on how amendments of A-level exams to reveal higher-order thinking skills were seen by teachers and students. Both groups welcomed the new examination policy and recognised similar positive backwash effects. Alkharusi (2010) and Monteiro et al. (2021) indicated that teachers perceived their formative practices more positively than their students did. Similar findings in the context of formative assessment were also reported in research by Cotton (2017) in the United States. Regarding formative assessment in the Netherlands, Pat-El et al. (2015) illuminated a substantial mismatch in perceptions between teachers and students, but Veugen et al. (2021) indicated a congruence between the assessment perceptions of students and teachers.

We build upon Monteiro et al. (2021), who emphasised a socio-constructivist viewpoint on assessment and concluded their results as follows (original italics): '…our results showed an inconsistency between teachers' conceptions and practices and more coherence between teachers' practices and students' conceptions. This allows us to think that these *teachers' assessment practices may in some way contribute to the way their pupils conceive the assessment process*' (p. 12). Therefore, we want to shed more light on assessment practices by employing integrated data on teachers and their students. How closely do assessment practices (teacher viewpoint) and their perceived usefulness for learning (student viewpoint) align?

STUDY AIM AND RESEARCH QUESTIONS

In this study, we examine the relations between teacher assessment practices and student perceptions of assessment. Our dataset derives from the first national evaluation of primary and lower secondary teacher assessment practices in the low-stakes assessment context of Finland.

First, we conduct a two-level analysis to answer our first research question (RQ):

(RQ1) Is the prevalence of assessment practices, as reported by teachers, correlated with students' perceived usefulness of these corresponding practices?

Second, we examine latent teacher profiles in terms of the diversity of their assessment practices. Moreover, we examine whether student perceptions of assessment differ in terms of these teacher subgroups. Our second RQ is divided into two parts and answered using latent class analysis:

(RQ2.1) What kinds of teacher profiles can be identified based on the diversity of assessment practices they report using?

(RQ2.2) Do students' perceptions of assessment differ in terms of their teachers' class?

Our study provides various insights regarding the diversity of classroom assessment. First, our study is located in the context of Finnish primary and lower secondary education.

Finland is known for its low-stakes and decentralised assessment culture (Kumpulainen & Lankinen, 2016). In such a context, teachers' assessment menus could arguably be rather diverse. Second, the significance of our study derives from a multilevel approach that considers both teachers' and students' points of view (Pat-El et al., 2015). Third, our study looks at a range of assessment methods simultaneously by relying on an 'assessment menu approach'.

METHODS

Context of the study

Our study concerns Finland's basic education, i.e., primary education (grades 1–6; ages 7–12 years) and lower secondary education (grades 7–9; ages 13–15 years). Assessment in Finland is characterised by a lack of standardised measurement (e.g., no national tests) and control (Vainikainen & Harju-Luukkainen, 2020; see Kumpulainen & Lankinen, 2016, for details). Assessment is low-stakes and based on local subject-specific curricular objectives and assessment criteria in two phases of basic education (grades 1–9; FNAE, 2020). The trust in teachers' professional autonomy to decide on assessment methods and practices is based on high-level pre-service teacher education in research-based universities.

However, despite teacher autonomy and the low-stakes assessment culture, recent studies have noted that closed-book examinations seem to be overemphasised in Finnish compulsory education (Atjonen et al., 2019; Mäkipää & Ouakrim-Soivio, 2019; Pollari, 2017). This may be due to the socio-historical role of examinations in the Finnish educational system: both students and teachers might be accustomed to examinations being the 'default' practice in assessment even though there are no national high-stakes tests at this educational level. The low-stakes assessment culture can be seen in the relative lack of study-related stress and anxiety. For example, an Organisation for Economic Co-operation and Development (OECD, 2015) report noted that Finnish students score extremely low when it comes to school- and test-related stress and anxiety.

Data collection and participants

The questionnaire used in this study was developed by experts in the Finnish Education Evaluation Centre (FINEEC) for national evaluation. Separate surveys were designed for students and teachers. Both surveys included a main topic of assessment practices, with several items formulated to fit the needs of respondent groups. We used data regarding the following seven assessment practices: closed-book examination, open-book examination, portfolio, written assignment or essay, self-assessment, peer-assessment and assessment discussions between students and teachers. Teachers rated the frequency, namely, how often each of these practices was used (5-point scale 1 = never to 5 = always). Students rated the extent to which they thought each assessment practice supported their learning (5-point scale 1 = 'Does not help at all' to 5 = 'Helps very much'). To answer RQ2.2, two sets of questions were used to map student perceptions of assessment (Tables 6 and 7):

- Perceived formative assessment actions (eigh items).
- Perceived pressures related to assessment (four items).

The dataset was collected through a questionnaire in January–February 2018 in a nationwide evaluation of Finnish assessment and feedback practices (Atjonen et al., 2019).

The dataset was collected by FINEEC. This was the first time such a sample-based, large-scale dataset on assessment has been collected in Finland. The data were collected by first sampling randomly from all Finnish- and Swedish-speaking (the two official languages of Finland) primary and lower secondary schools, excluding specialist schools. This collection of 289 schools was sampled to be representative of all the Finnish schools with their demographic aspects (language, area, school type). Second, teachers were sampled so that they evenly represented all school subjects, as categorised into six groups (mathematics and science, foreign languages, Finnish language and literature, expressive arts, humanities and counselling). In the third phase, students were sampled based on the disciplines their teacher taught.

Datasets of online surveys were first saved separately, including 1709 teachers and 5001 students. Secondly, teachers and students were paired by means of a school identification number. In this phase, data went missing as not all students could be connected to their teacher due to missing school identification numbers. From the broader dataset, we chose teachers and students of sixth and ninth grades that end the primary (grades 1–6) and lower secondary (grades 7–9) levels of Finnish compulsory comprehensive school (excluding general upper secondary education). Based on the initial Finnish report (Atjonen et al., 2019), we knew that the distribution of teachers' assessment methods was rather similar at both primary and lower secondary levels, which is why we included both grade levels in the dataset without a comparative interest. Moreover, teachers at both grade levels are guided by the same general assessment guidelines (FNAE, 2020) as well as national legislation.

Finally, our two-level dataset consisted of 187 teachers and their 2370 students. Of these 187 teachers, 74 were sixth-grade teachers and 113 ninth-grade teachers.

Data analysis

To answer RQ1, we used a two-level modelling in MPlus (Muthén & Muthén, 1998–2017). The two-level analysis acknowledges the hierarchical nature of data (de Leeuw & Meijer, 2008). This analysis was used to test correlations between teachers' self-reported use of assessment practices and students' perceived usefulness of these practices between levels (while considering the variation of teacher responses). In the analysis, the teacher dataset and the student dataset constituted two levels. In our model, we had two levels: the differences *within* levels (the variance within one classroom; Level 1) and the differences *between* levels (the variance between different classrooms; Level 2). The students' responses could vary at both Levels 1 and 2, whereas the teachers' responses could only vary at Level 2. Intraclass correlation was calculated for student variables. Intraclass correlation is the proportion of variance associated to between level (between teachers) and student variables. The correlations were calculated at Level 2 (between levels).

RQ2.1 was addressed by conducting an LCA (Latent Class Analysis) to identify homogeneous latent subgroups of teachers according to their assessment practices. LCA offers a person-oriented analysis to classify individuals into homogenous subgroups by their latent, underlying classes (Collins & Lanza, 2010). The number of classes is presumed to be unknown. Unlike many other profiling and clustering methods (e.g., hierarchical cluster analysis), LCA provides fit indexes to test the various potential class solutions. We used the teacher-reported prevalence for seven different assessment practices as the profiling variables in LCA (see Table 1). The LCA was conducted for four latent class solutions. The analysis was conducted in MPlus. We used two fitness indicators to evaluate the number of latent groups: the Bayesian information criterion (BIC) and bootstrapped likelihood ratio test (BLRT). The lower the BIC, the better the model fits the data. A significant BLRT result rejects the null hypothesis of $k-1$ latent group solution and accepts the alternative hypothesis

TABLE 1 Description of the dataset.

Teacher-reported prevalence	Mean	SD	Student-reported perceived usefulness	Mean	SD
Closed-book examination	4.13	1.10	Closed-book examination	3.16	1.11
Open-book examination	2.23	1.01	Open-book examination	3.05	1.24
Portfolio	1.92	1.06	Portfolio	2.35	1.23
Written assignment, essays	3.01	1.42	Written assignment, essays	2.76	1.23
Self-assessment	3.48	1.01	Self-assessment	2.56	1.12
Peer-assessment	2.58	1.00	Peer-assessment	2.53	1.12
Assessment discussions	2.23	1.04	Assessment discussions	2.52	1.26

Note: All variables were measured on a 1–5 Likert scale.

k latent group solution. Average posterior probability (AvePP) shows the distinctiveness between latent classes. AvePP is a value between 0 and 1; the closer the value is to 1, the more clearly the groups stand out. The latent class profiles were calculated by multiplying the probability of a given subclass by its value (categories are numbered 1, 2, 3, 4, 5). Mean differences in student variables between latent classes were tested with the BCH test implemented in MPlus. BCH weights were saved to a file and then used to calculate means and standard deviations and to test the differences between classes (RQ2.2). While testing students' variables (RQ2.2), type COMPLEX was used to correct bias due to the intraclass correlation. If the Wald test was significant, the pairwise comparison of classes was tested with the help of the new parameters in Mplus.

FINDINGS AND DISCUSSION

(RQ1) The prevalence of assessment practices and the correlations with students' perceived usefulness of assessment

The means and standard deviations of the teacher-reported prevalence of assessment practices and student-reported usefulness of these same practices are reported in Table 1. Closed-book examinations were the most commonly used assessment practice, and students considered this practice the most useful for learning. In comparison, other assessment practices were used less often.

The correlations in Table 2 indicate that teachers' practices and students' views of usefulness aligned more strongly between essays and exams than they did for, for example, self- and peer-assessment. The intraclass correlations for student variables varied between 0.05 and 0.19 (Table 2). These findings emphasise that broadly, there is a connection—albeit not very strong—in Finland between a teacher's prevalence of using individual assessment practices and student perceptions of the corresponding practices. The most commonly used assessment practices are also seen by students as the most useful for learning.

These findings complement earlier investigations in Finnish language education (e.g., Mäkipää, 2020) and mathematics education (Nieminen & Atjonen, 2022). Despite its low-stakes assessment context, Finnish basic education relies largely on closed-book examinations. Other practices, such as portfolios and peer-assessments, were reported to be used less. Table 2 shows that when alternative assessment methods were used more often, students generally perceived them as more helpful for their learning; however, the correlations varied. It can be hypothesised (yet not deducted since our analysis is correlational) that

TABLE 2 Correlations between teacher assessment practices and their usefulness as perceived by students.

Assessment practice	Between class correlation	Intraclass correlation
Closed-book examination	0.61*	0.13*
Open-book examination	0.49*	0.11*
Portfolio	0.40*	0.13*
Written assignment, essays	0.70*	0.19*
Self-assessment	0.26*	0.05**
Peer-assessment	0.36**	0.05*
Assessment discussions	0.46*	0.13*

*$p < 0.001$; **$p < 0.01$.

TABLE 3 Fit indices for four class solutions.

	BIC	p_{BLRT}	AvePP
1 class	3585.42	–	–
2 classes	3472.89	<0.001	0.92
3 classes	3547.33	<0.001	0.91
4 classes	3642.86	0.667	0.99

exposing students to diverse assessment practices may enhance students' understanding of how these practices may promote their learning.

(RQ2.1) Teacher classes

Next, we will present and discuss our findings in relation to the various subclasses of teachers when it comes to their 'assessment menu'. With LCA, we tested four class solutions as presented in Table 3.

According to the BIC index, the solution of two latent classes fits the data best. The BLRT test was statistically significant while testing the 2-class solution against the 1-class solution, as was the 3-class solution against the 2-class solution. However, the BLRT index was nonsignificant while testing the 4-class solution against the 3-class solution, which is why we singled this solution out. According to the AvePP values, the latent teacher classes were clearly distinctive.

We chose the 3-class solution since it differentiates between three subgroups, in opposition to providing a purely dichotomous solution with two potentially heterogeneous classes (Figure 1 and Table 4). This suited our RQ which aimed to understand qualitatively different teacher subgroups when it comes to teachers' assessment practices. This way, we could also better understand potential extreme profiles as with the two-class solution. Earlier research supports the emergence of various teacher profiles with respect to assessment (e.g., Paakkari et al., 2022; Veldhuis & van den Heuvel-Panhuizen, 2014). Closed-book exams were used widely in the three classes, with self-assessment being the second-most used practice in the classes.

Figure 1 shows that the three classes had few qualitative differences. Instead, they represented 'high', 'medium' and 'low' groups, following a rather similar distribution profile (see also Table 4). This is in contrast to earlier studies on teacher profiles in assessment which
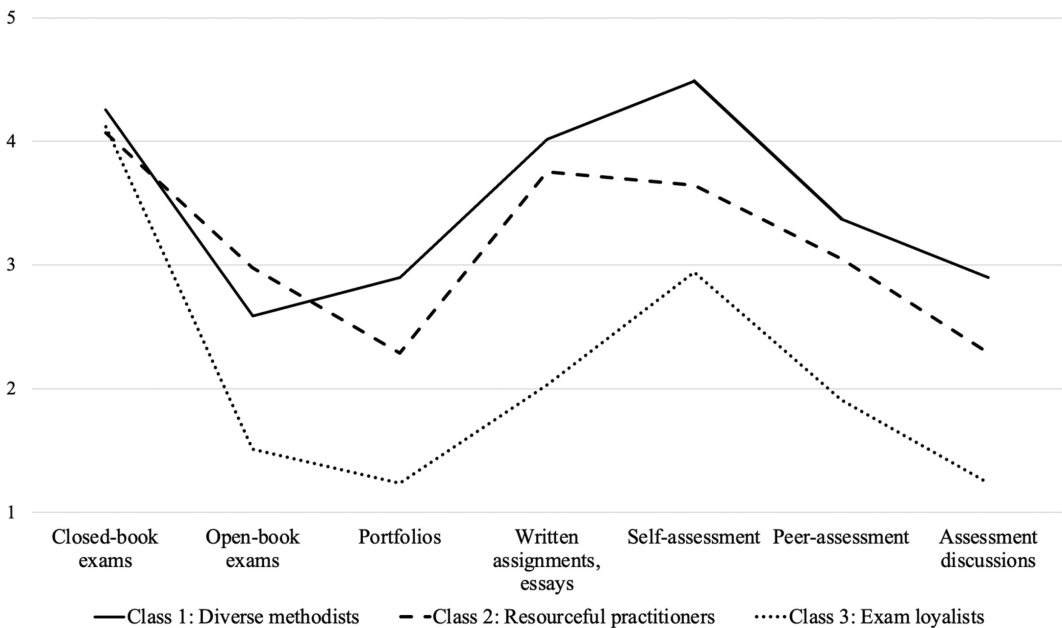
FIGURE 1   The three teacher classes.

identified more qualitatively nuanced clusters and profiles (e.g.Paakkari et al., 2022; Picos et al., 2013; Veldhuis & Van den Heuvel-Panhuizen, 2014).

While naming these teacher classes, we did not want to assign normative names to the groups (e.g., 'problematic assessors' as used by Paakkari et al., 2022). Without background data on the teachers' intentions or motivations, we avoided ascribing meaning to the class names beyond the methods used (cf. 'enthusiastic teachers' in Veldhuis & Van den Heuvel-Panhuizen, 2014). Finally, we drew on an exploratory, data-driven approach by naming the classes as follows:

- *Class* 1: Diverse methodists: teachers who used the widest diversity of assessment practices (*35 teachers with 370 students*)
- *Class* 2: Resourceful practitioners: teachers who quite often use closed-book exams but also many other practices (*69 teachers with 943 students*)
- *Class* 3: Exam loyalists: teachers who reported using mainly closed-book exams (*83 teachers with 1057 students*)

We tested the distribution of the three teacher classes (chi-square test of independence) in terms of disciplinary background and grade level. First, a statistically significant difference ($\chi^2(8, N=187)=55.93, p<0.001$) with a medium effect size (Cramer's $V=0.39$) was identified for their disciplinary background. Teachers from the natural sciences were overrepresented in Class 3 (exam-oriented), and teachers from Finnish/Swedish language and literature were overrepresented in Class 2 ('Resourceful practitioners'). There was no statistically significant difference ($\chi^2(2, N=187)=5.03, p=0.08$) in terms of the grade level taught (grade six or nine).

The most diverse teacher class was also the smallest one, which again aligns with earlier studies noting that assessment in the Finnish education system is rather test-focused despite low-stakes assessment cultures and policies. This is exemplified by the fact that closed-book exams were used commonly in each of the three classes. As noted, the largest

class had an overrepresentation of teachers from the natural sciences, which might reflect the testing-oriented culture of these disciplines (Iannone & Simpson, 2022; Martínez-Sierra et al., 2016; Watt, 2005).

(RQ2.2) Student perceptions of assessment by teacher class

The next step in our analysis was to compare the students' perceived usefulness of assessment according to the three teacher classes. As noted in Figure 2 and Table 5, the differences were minor. The differences between students whose teachers belonged to Classes 1 and 2 ('diverse methodists' and 'resourceful practitioners') were almost non-existent. In some assessment practices (open-book exams, portfolios, written assignments and essays), we identified statistically significant differences, but the effect sizes were small $(0.02 < \eta^2 < 0.06)$.

Finally, regarding RQ2.2, we report the students' reported (1) formative assessment activities (eight items) and (2) pressures of assessment (four items). There were no statistically significant differences in terms of the perceived formative assessment activities (Table 6). Regarding student perceptions of the pressures of assessment, statistically significant differences in terms of the teacher's subclass (Table 7) were identified. There was a trend that students whose teachers were classified as 'diverse methodists' reported higher levels of perceived pressure. This trend was seen in three of the four items.

Regardless of teacher class, students reported a relatively high prevalence of formative assessment activities (Table 6). Given that teachers often perceive formative assessment practices in a more auspicious light than students (see, e.g., Cotton, 2017; Monteiro

TABLE 4 The teacher classes.

	Class 1[a]		Class 2[b]		Class 3[c]		Total		ANOVA		Post-hoc testing
	M	SD	*M*	SD	*M*	SD	*M*	SD	*F*(2, 183)	η^2	Tukey HSD
Closed-book examination	4.35	0.98	4.04	0.53	4.11	1.44	4.13	1.10	0.93	–	–
Open-book examination	2.53	0.92	2.97	0.67	1.49	0.71	2.23	1.01	76.82*	0.47	2 > 1** 2 > 3* 1 > 3*
Portfolios	2.91	1.23	2.28	0.88	1.21	0.50	1.92	1.06	59.96*	0.41	1 > 2 > 3*
Written assignments, essays	4.09	1.38	3.71	0.67	1.99	1.21	3.01	1.42	67.56*	0.43	1 > 3* 2 > 3*
Self-assessment	4.44	0.66	3.66	0.73	2.94	0.99	3.48	1.01	40.13*	0.31	1 > 2 > 3*
Peer-assessment	3.33	0.82	3.06	0.71	1.89	0.80	2.58	1.00	61.61*	0.41	1 > 3* 2 > 3*
Assessment discussions	2.18	1.21	2.36	0.79	2.15	1.13	2.23	1.04	0.81	–	–

Note: All the variables met the assumption of homogeneity of variance for the post-hoc tests.

[a]Diverse methodists.

[b]Resourceful practitioners.

[c]Exam loyalists.

*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$

et al., 2021), the assessment perceptions of Finnish students seem rather positive. Moreover, all students reported relatively low levels of pressure related to assessment (Table 7). One explanation for these findings is the overall low-stakes culture of Finnish classroom assessment. This is in alignment with the OECD report (2015) report noting that Finnish students score very low when it comes to school- and test-related stress and anxiety.

GENERAL DISCUSSION OF THE FINDINGS

The first part of our findings (RQ1) showed that teacher assessment practices and students' perceived usefulness of assessment were correlated at a broad level. However, a further LCA for teachers did not reveal differences in their corresponding student populations (RQ2). While it was possible to differentiate three separate teacher classes based on their 'assessment menus', the assessment perceptions of the students of these teachers were almost identical. This might result from the domination of examinations in all teacher classes. The teacher classes were not qualitatively different but represented 'high', 'medium' and 'low' groups. In other terms, the teachers' assessment practices did not explain student perceptions of assessment, but the three student subgroups provided a rather homogeneous view of assessment. Thus, our study produced few statistically significant findings. This is an evocative finding in itself as it emphasises the rather homogeneous assessment culture in Finnish primary and lower secondary education.

Interestingly, there was a trend towards greater pressure perceived by those students whose teachers were labeled as 'diverse methodologists' (Table 7). While the differences were small, this phenomenon warrants further investigation. In high-stakes assessment cultures, diverse assessment menus may be perceived by students as hindering their focus on achieving high scores in examinations (Guo & Yan, 2019). Yet, even in low-stakes assessment cultures, students may have conservative and negative attitudes towards formative
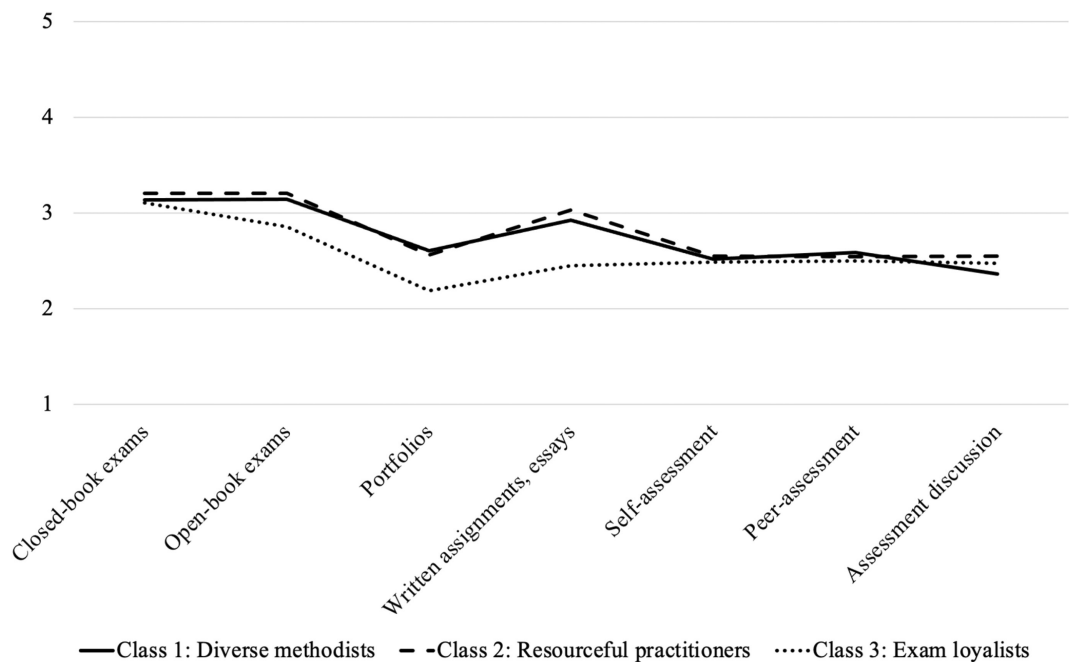

FIGURE 2   Students' perceived usefulness of assessment according to teacher class.

TABLE 5 Students' perceived usefulness of assessment practices according to teacher class.

Perceived usefulness of…	Class 1[a]		Class 2[b]		Class 3[c]		Total		Wald test[d]	Pairwise comparisons
	M	SD	*M*	SD	*M*	SD	*M*	SD		
Closed-book exam	3.21	1.04	3.24	1.04	3.06	1.20	3.16	1.12	4.93	–
Open-book examination	3.15	1.23	3.20	1.14	2.88	1.31	3.05	1.24	12.77***	1 > 3*** 2 > 3*
Portfolios	2.44	1.28	2.53	1.18	2.15	1.23	2.35	1.23	20.59*	1 > 3*** 2 > 3*
Written assignments, essays	2.91	1.16	3.06	1.13	2.42	1.25	2.76	1.23	48.93*	1, 2 > 3*
Self-assessment	2.58	1.10	2.62	1.06	2.49	1.17	2.56	1.12	3.81	–
Peer-assessment	2.57	1.09	2.58	1.06	2.46	1.18	2.53	1.12	3.81	–
Assessment discussions	2.45	1.22	2.62	1.22	2.45	1.31	2.52	1.26	3.06	–

[a]Diverse methodists.
[b]Resourceful practitioners.
[c]Exam loyalists.
[d]Degrees of freedom = 2.
*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$.

assessment practices (e.g., Dhindsa et al., 2007). Unfortunately, our study provides no explanations for such findings. Earlier research has noted that diverse assessment menus may be challenging for students, particularly when the assessment workload is perceived as heavy (e.g., Gijbels & Dochy, 2006). Moreover, students might be unable to make use of alternative assessment practices due to their lack of assessment literacies (Hannigan et al., 2022).

Our findings indicate that simply widening the menu of assessment may not necessarily lead to positive and productive perceptions of assessment. We hypothesise that this may particularly be the case if student assessment literacy has not been trained and developed; yet this is an avenue for future research. Another possibility is that the teachers who reported using many assessment practices *commonly* did not, in fact, use these practices *adequately*. A diverse assessment menu is thus not an answer in itself but needs to be carefully designed to match the learning outcomes and pedagogical practices.

CONCLUDING REMARKS, LIMITATIONS AND IMPLICATIONS

We examined the interrelations between teachers' assessment menus and students' perceptions of assessment. A multilevel analysis of 187 Finnish primary and lower secondary teachers and their 2370 students showed rather strong yet varying correlations between the diversity of assessment and students' perceived usefulness of assessment (RQ1). Overall, the students largely perceived assessment as not very helpful for learning (Figure 2). Moreover, LCA revealed that the teachers could be divided into three subgroups with varying assessment menus (RQ2.1), yet these subgroups showed few qualitative differences (see Figure 1). Closed-book examinations were used widely by teachers in all three subgroups, and the class 'Exam loyalists' was the largest subgroup. There were only minor differences in students' perceptions of assessment with respect to their teachers' classes in LCA (RQ2.2). While the multilevel analysis emphasised a connection between teacher

TABLE 6 Students' perceived formative assessment activities according to teacher class.

The teacher has…	Class 1[a]		Class 2[b]		Class 3[c]		Total		Wald test[d]	Pairwise comparisons
	M	SD	M	SD	M	SD	M	SD		
Helped me revise my mistakes and learn from them	3.52	1.09	3.52	1.04	3.58	1.09	3.55	1.07	0.12	–
Encouraged me to try harder through assessment	3.42	1.04	3.36	1.11	3.43	1.18	3.40	1.13	0.69	–
Told me how to progress in my studies	3.37	1.08	3.42	1.08	3.39	1.12	3.40	1.10	0.32	–
Helped me set my own goals	3.32	0.97	3.35	1.08	3.33	1.09	3.34	1.07	0.90	–
Helped me develop my ways of learning	3.23	1.03	3.27	1.04	3.31	1.10	3.28	1.07	0.53	–
Informed me about my strengths	3.11	1.19	3.19	1.24	3.10	1.22	3.14	1.22	1.06	–
Guided me to assess if I have reached my goals	3.14	1.04	3.14	1.05	3.09	1.07	3.12	1.06	1.17	–
Told me how to prepare for assessment	2.90	1.17	2.99	1.11	2.93	1.12	2.95	1.15	0.99	–

[a]Diverse methodists.

[b]Resourceful practitioners.

[c]Exam loyalists.

[d]Degrees of freedom = 2.

*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$.

assessment practices and student perceptions, the person-oriented analysis implies that the interplay may be rather complex in reality. To sum up, based on our findings, classroom assessment is not very diverse in Finland. At the same time, students do not find assessment as something very meaningful for their learning, yet students also do largely not report stress or pressure about assessment.

This study adds significant knowledge to our understanding of classroom assessment in the Finnish education system (Kumpulainen & Lankinen, 2016; Thrupp et al., 2023). Based on our findings, Finland seems to be a context where the low stakes of assessment and the lack of high-stakes testing have not led to diverse assessment practices being used in schools. Our findings clash with the ethos of the national curriculum that emphasises diverse and formative assessment practices (FNAE, 2014). These findings contest the general idea that assessment would necessarily be more diverse in low-stakes assessment cultures. Why might assessment not be diverse in Finland, and what might explain the students' perceptions of assessment in this context? While our data provides no answers to these questions, we hypothesise that this may be due to the history of the educational system in Finland. While Finland is commonly portrayed as a 'PISA miracle' with up-to-date

TABLE 7 Students' perceived pressures of assessment according to teacher class.

	Class 1[a]		Class 2[b]		Class 3[c]		Total		Wald test[d]	Pairwise comparisons
	M	SD	M	SD	M	SD	M	SD		
My teacher is too demanding or strict in assessment	2.68	1.17	2.48	1.11	2.32	1.10	2.44	1.12	9.81**	1 > 3**
Assessment situations frighten and distress me beforehand	2.48	1.28	2.41	1.26	2.30	1.29	2.37	1.28	3.29	–
There are too many examinations and assessed tasks	2.87	1.27	2.79	1.20	2.60	1.24	2.72	1.23	6.75***	1 > 3***
The teacher gives me too much negative feedback	2.22	1.18	2.06	1.11	1.89	1.10	2.01	1.12	13.40**	1 > 3** 2 > 3***

[a]Diverse methodists.
[b]Resourceful practitioners.
[c]Exam loyalists.
[d]Degrees of freedom = 2.
*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$.

pedagogies, in fact, the educational system still largely relies on a rather authoritarian mode of teaching that draws on pedagogical conservatism (see Simola, 2005). In our earlier study (Nieminen & Atjonen, 2022), we noted that when this conservatism collides with the low stakes of assessment, assessment might entail a 'culture of compliance'. In such cultures, examinations may dominate assessment due to their socio-historical weight. However, examinations may not cause similar stress and anxiety for students as they might in high-stakes assessment cultures. As our findings indicate, within 'cultures of compliance', it is possible that diverse assessment practices cause students stress. This is something that future studies on implementing student-centred assessment practices in low-stakes assessment cultures should take into account.

We conclude that future research must better understand the (lack of) diversity of classroom assessment, and the student perceptions that follow, in low-stakes cultures. Research literature has listed various barriers to diversifying assessment in high-stakes cultures, such as teachers' lack of assessment literacy (Siegel & Wissehr, 2011), the washback effect of high-stakes testing (Chan, 2020) and lack of time and resources (O'Neill & Padden, 2022). How similar ideas operate in low-stakes cultures warrants further investigation. Our findings indicate that in the low-stakes assessment culture of Finland, the assessment menu teachers use is not directly related to student perceptions of assessment; instead, students seem to be rather homogeneous in their perceptions. In such contexts, diversifying assessment may not provide a magic bullet to improving student perceptions of assessment and its usefulness.

One possible way to interpret our findings relates to the deep-rooted socio-cultural and -political assumptions of *what assessment is* and what kinds of norms and ideologies guide the questions of *what assessment should be* in a given context. Some culturally shared beliefs about assessment, such as the idea of exams as the 'gold standard' of assessment (Richardson, 2022), may be so deeply ingrained in our societies that they are shared largely by the wider public, including parents and policy-makers (Nieminen et al., 2021). We propose that future research could unpack the practices and values of assessment by framing them in their wider socio-historical and -political contexts (see Simola, 2005).

We note some important limitations in our work. First, our exploratory study drew on a national evaluation survey by FINEEC. The lack of validated research instruments is a profound limitation of our work. Second, the survey was conducted using the same items for both sixth and ninth graders, who may have perceived the items differently despite the scrutiny in formulating the items with age-sensitive language and the rigorous piloting of the questionnaire. Third, we do not know *how* the teachers used the various assessment practices, as the survey only mapped out their prevalence. Importantly, we do not know whether the teachers had implemented the assessment practices summatively or formatively. Fourth, our sample of 189 teachers was relatively small, meaning that we could not identify latent groups that might have been identifiable from a bigger sample. The small size of each of the three teacher classes reduces the statistical power of further comparative analyses. However, even with this sample size, we were able to identify three clearly distinctive subgroups of teachers. Fifth, a significant amount of data went missing due to absent school codes while integrating teacher and student datasets. Finally, and perhaps most importantly, we emphasise that our study concerns correlations, not causations. This provides an important caveat for future research in understanding (i) how exactly teachers' assessment menus influence student perceptions of assessment and (ii) how students' perceptions might, in turn, Ba limit teachers from diversifying their assessment practices.

Based on our findings, we recommend further support for both in- and pre-service teachers for diversifying their assessment menus in ways that are pedagogically meaningful. The social, cultural and historical role of examinations in various assessment cultures might need to be carefully scrutinised in such work. However, it is not only teachers' assessment literacy that requires attention but that of students' as well. Ideally, a diverse assessment menu would promote students' assessment literacies by providing them with multiple forms of information about their skills and knowledge (Hannigan et al., 2022). By developing students' assessment literacy, teachers may create more fruitful ground for diverse assessment menus to flourish. We call for a shared responsibility of assessment between students and teachers through shared assessment literacies (see Carless & Winstone, 2022). Student–teacher partnerships and dialogues in assessment design may offer valuable practical solutions for calibration assessment practices and perceptions with the multiple stakeholders in assessment.

Finally, we emphasise that fostering the purpose of assessment to promote learning (rather than only measure it) is a significant challenge both in low- and high-stakes assessment cultures. Diversifying assessment is not only a technical demand for a broader menu of methods but instead, this quest is deeply connected with the societal role of assessment: how assessment is valued, what 'good assessment' is perceived to be and how assessment is implemented in practice. We welcome future research that unpacks student and teacher perspectives of the diversity of assessment through situated approaches that shed light on such values and ideologies.

DATA AVAILABILITY STATEMENT
The data is not publicly available since it is a national evaluation dataset. The data that support the findings of this study are available from The Finnish Education Evaluation Centre (FINEEC). Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of The Finnish Education Evaluation Centre (FINEEC).

ETHICS STATEMENT

The project has received an ethics approval.

GEOLOCATION INFORMATION

The data were collected from Finland.

ORCID

Juuso Henrik Nieminen ⓘ https://orcid.org/0000-0003-3087-8933
Päivi Atjonen ⓘ https://orcid.org/0000-0002-9864-9339
Asko Tolvanen ⓘ https://orcid.org/0000-0001-6430-8897

TWITTER

Juuso Henrik Nieminen 🐦 JuusoNieminen

REFERENCES

Alkharusi, H. (2010). Teachers' assessment practices and students' perceptions of the classroom assessment environment. *World Journal on Educational Technology*, *2*(1), 27–41.

Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., & Immonen, J. (2019). *"Että tietää missä on menossa": oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa ["So that we know where we stand" assessment of learning and competence in basic education and general upper secondary education]*. The Publications of the Finnish Education Evaluation Centre no. 7.

Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systemic curricular reform? *Journal of Curriculum Studies*, *32*(5), 623–650.

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, *13*(1), 117–140.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, *18*(4), 451–469.

Bourke, R., & Mentis, M. (2014). An assessment framework for inclusive education: Integrating assessment approaches. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 384–397.

Brooks, C., Huang, Y., Hattie, J., Carroll, A., & Burton, R. (2019). What is my next step? School students' perceptions of feedback. *Frontiers in Education*, *4*, 96.

Brown, G. T. (2022). Student conceptions of assessment: Regulatory responses to our practices. *ECNU Review of Education*, *5*(1), 116–139.

Carless, D., & Winstone, N. (2022). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, *28*(1), 150–163.

Chan, K. L. R. (2020). Washback in education: A critical review and its implications for language teachers. *Journal of Foreign Language Education and Technology*, *5*(1), 108–124.

Chan, K. T., & Tan, K. (2022). How teachers experience assessment tension and its effect on formative assessment practices. *Educational Research for Policy and Practice*, *21*(3), 447–464.

Chu, H. C. (2014). Potential negative effects of mobile learning on students' learning achievement and cognitive load—A format assessment perspective. *Journal of Educational Technology & Society*, *17*(1), 332–344.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons.

Cotton, D. (2017). Teachers' use of formative assessment. *Delta Kappa Gamma Bulletin*, *83*(3), 39.

Daly, A. L., Baird, J. A., Chamberlain, S., & Meadows, M. (2012). Assessment reform: Students' and teachers' responses to the introduction of stretch and challenge at A-level. *Curriculum Journal*, *23*(2), 139–155.

de Leeuw, J., & Meijer, E. (Eds.). (2008). *Handbook of multilevel analysis*. Springer-Verlag.

DeLuca, C., & Lam, C. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly*, *41*(3), 3–24.

DeLuca, C., Rickey, N., & Coombas, A. (2021). Exploring assessment across cultures: Teachers' approaches to assessment in the U.S., China, and Canada. *Cogent Education*, *8*(1), 1921903. https://doi.org/10.1080/2331186X.2021.1921903

Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education*, *29*(10), 1261–1280.

FNAE (Finnish National Agency for Education). (2020). Assessment of pupils' learning and competence in basic education. Changes to the "National core curriculum for basic education 2014". https://www.oph.fi/fi/koulutus-ja-tutkinnot/arviointiluku-6-oppimisen-ja-osaamisen-arviointi

FNAE (Finnish National Agency for Education). (2014). *The national core curriculum for basic education 2014* (p. 5). Finnish National Agency for Education. Publications 2016.

Fulmer, G. W., Lee, I. C., & Tan, K. H. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 475–494.

Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, *32*(4), 399–409.

Graham, L. J., Tancredi, H., Willis, J., & McGraw, K. (2018). Designing out barriers to student access and participation in secondary school assessment. *The Australian Educational Researcher*, *45*(1), 103–124.

Guo, W. Y., & Yan, Z. (2019). Formative and summative assessment in Hong Kong primary schools: students' attitudes matter. *Assessment in Education: Principles, Policy & Practice*, *26*(6), 675–699.

Hannigan, C., Alonzo, D., & Oo, C. Z. (2022). Student assessment literacy: Indicators and domains from the literature. *Assessment in Education: Principles, Policy & Practice*, *29*(4), 482–504.

Harris, L. R., & Brown, G. T. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 365–381.

Iannone, P., & Simpson, A. (2022). How we assess mathematics degrees: The summative assessment diet a decade on. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *41*(1), 22–31.

Johansson, E., Kanapathippillai, S., Khan, A., & Dellaportas, S. (2022). Formative assessment in accounting: Student perceptions and implications of continuous assessment. *Accounting Education*, *32*(6), 597–625. https://doi.org/10.1080/09639284.2022.2091411

Kim, N., & Son, Y. (2021). Multilevel latent profile analysis of Korean middle school student perceptions of teaching methods. *Asia Pacific Education Review*, *24*, 41–55.

Kippers, W., Wolterinck, C., Schildkamp, K., Poortman, C., & Visscher, A. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199–213.

Kumpulainen, K., & Lankinen, T. (2016). Striving for educational equity and excellence: Evaluation and assessment in Finnish basic education. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp. 71–82). Sense.

Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2024). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, *47*(3), 529–547. https://doi.org/10.1080/02619768.2022.2058927

Lin, L., Li, G. Y., & Guo, X. (2024). Pre-service Chinese language teachers' conceptions of assessment: A person-centered perspective. *Language Teaching Research*, *28*(1), 273–295.

Liu, J. E., Lo, Y. Y., & Xin, J. J. (2023). CLIL teacher assessment literacy: A scoping review. *Teaching and Teacher Education*, *129*, 104150.

Mäkipää, T. (2020). Feedback practices in language classes in Finnish general upper secondary schools. *Apples-Journal of Applied Language Studies*, *14*(1), 103–123.

Mäkipää, T., & Ouakrim-Soivio, N. (2019). Finnish upper secondary school students' perceptions of their teachers' assessment practices. *Journal of Teaching and Learning*, *13*(2), 23–42.

Martínez-Sierra, G., Valle-Zequeida, M. E., Miranda-Tirado, M., & Dolores-Flores, C. (2016). Social representations of high school students about mathematics assessment. *Canadian Journal of Science, Mathematics and Technology Education*, *16*, 247–258.

Monteiro, V., Mata, L. & & Nóbrega Santos, N. (2021) Assessment conceptions and practices: Perspectives of primary school teachers and students. *Frontiers in Education*, *6*, 631185. https://doi.org/10.3389/feduc.2021.631185

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nieminen, J. H., & Atjonen, P. (2022). The assessment culture of mathematics in Finland: A student perspective. *Research in Mathematics Education*, *25*(2), 243–262.

Nieminen, J. H., & Lahdenperä, J. (2024). Assessment and epistemic (in) justice: How assessment produces knowledge and knowers. *Teaching in Higher Education*, *29*(1), 300–317.

Nieminen, J. H., Atjonen, P., & Remesal, A. (2021). Parents' beliefs about assessment: A conceptual framework and findings from Finnish basic education. *Studies in Educational Evaluation*, *71*, 101097.

O'Neill, G., & Padden, L. (2022). Diversifying assessment methods: Barriers, benefits and enablers. *Innovations in Education and Teaching International*, *59*(4), 398–409.

OECD. (2015). *PISA 2015 results* (Vol. *III*, pp. 1–528). Students' Well-Being. https://www.oecd-ilibrary.org/docserver/9789264273856-en.pdf?expires=1540817348&id=id&accname=guest&checksum=B1F49F6578A71D8EAF15F19F5FDE8541

Paakkari, O., Paakkari, L., Haapala, H., & Hirvensalo, M. (2022). Health education teachers' assessment conceptions and practices: Identifying assessment profiles. *Educational Assessment*, *27*(3), 285–299.

Pastore, S., Manuti, A., & Fausta Scardigno, A. (2019). Formative assessment and teaching practice: The point of view of Italian teachers. *European Journal of Teacher Education*, *42*(3), 359–374.

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2015). Multilevel predictors of differing perceptions of assessment for learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 282–298.

Picos, A. P., López-Pastor, V. M., & Barba, J. J. (2013). Clusters-models of teachers about students' assessment in teacher education. *Estudios Sobre Educación*, *24*, 173.

Picos, P. A., & Lopez-Pastor, M. V. (2013). Do what I say, not what I do: Student assessment systems in initial teacher education. *Revista De Educacion*, *361*, 279–305.

Pollari, P. (2017). The power of assessment: What (dis) empowers students in their EFL assessment in a Finnish upper secondary school? *Apples-Journal of Applied Language Studies*, *11*(2), 147–175

Ratnam-Lim, C. T. L., & Tan, K. H. K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 61–78.

Richardson, M. (2022). *Rebuilding public confidence in educational assessment*. UCL Press.

Shepard, L. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066–1101). AERA.

Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. *The Phi Delta Kappan*, *69*(1), 38–44.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371–391.

Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, *41*(4), 455–470.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Strijbos, J. W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the feedback perceptions questionnaire. *Studies in Educational Evaluation*, *68*, 100980.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, *30*(4), 325–341.

Thrupp, M., Seppänen, P., Kauko, J., & Kosunen, S. (2023). *Finland's famous education system: Unvarnished insights into Finnish schooling*. Springer Nature.

Vainikainen, M. P., & Harju-Luukkainen, H. (2020). Educational assessment in Finland. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring student achievement in the 21st century* (pp. 131–142). Springer.

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*(2), 345–373.

Vattøy, K. D., Gamlem, S. M., Kobberstad, L. R., & Rogne, W. M. (2022). Students' experiences of assessment and feedback engagement in digital contexts: A mixed-methods case study in upper secondary school. *Education Inquiry*, 1–22. https://doi.org/10.1080/20004508.2022.2122202

Veldhuis, M., & van den Heuvel-Panhuizen, M. (2014). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE*, *9*(1), 1–11.

Veugen, M., Gulikers, J., & den Brok, P. (2021). We agree on what we see: Teacher and student perceptions of formative assessment practice. *Studies in Educational Evaluation*, *70*, 101027. https://doi.org/10.1016/j.stueduc.2021.101027

Volante, L., DeLuca, C., Adie, L., Baker, E., Harju-Luukkainen, H., Heritage, M., Schneider, C., Stobart, G., Tan, K., & Wyatt-Smith, C. (2020). Synergy and tension between large-scale and classroom assessment: International trends. *Educational Measurement: Issues and Practice*, *39*(4), 21–29.

Warwick, P., Shaw, S., & Johnson, M. (2015). Assessment for learning in international contexts: Exploring shared and divergent dimensions in teacher values and practices. *The Curriculum Journal*, *26*(1), 39–69.

Watt, H. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics*, *58*, 21–44.

Winstone, N. E., & Carless, D. (2021). Who is feedback for? The influence of accountability and quality assurance agendas on the enactment of feedback processes. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 261–278.

Winterbottom, M., Taber, K. S., Brindley, S., Fisher, L. G., Finney, J., & Riga, F. (2008). Understanding differences in trainee teachers' values and practice in relation to assessment. *Teacher Development*, *12*(1), 15–35.

Wurf, G., & Povey, R. (2020). "They show how smart you are": A mixed methods study of primary students' perceptions of assessment tasks. *Issues in Educational Research*, *30*(3), 1162–1182.

Yan, Z., & Brown, G. T. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985.

Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the teacher formative assessment literacy scale. *Studies in Educational Evaluation*, *74*, 101183.

Yang, Y., & Hu, B. Y. (2019). Chinese preschool teachers' classroom instructional support quality and child-centered beliefs: A latent profile analysis. *Teaching and Teacher Education: An International Journal of Research and Studies*, *80*(1), 1–12.