



OPEN ACCESS

EDITED BY

Pernille Fiskerstrand,
Volda University College, Norway

REVIEWED BY

Maurizio Gentile,
Libera Università Maria SS. Assunta, Italy
Fauziah Sulaiman,
University Malaysia Sabah, Malaysia

*CORRESPONDENCE

Andrea Westphal
✉ andrea.westphal@uni-greifswald.de

RECEIVED 04 November 2024

ACCEPTED 25 March 2025

PUBLISHED 16 April 2025

CITATION

Westphal A, Hoferichter CJ and Vock M
(2025) What does it take for students to
value grades for oral participation?
Transparency is key.
Front. Educ. 10:1522695.
doi: 10.3389/feduc.2025.1522695

COPYRIGHT

© 2025 Westphal, Hoferichter and Vock. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

What does it take for students to value grades for oral participation? Transparency is key

Andrea Westphal^{1*}, Clara Josepha Hoferichter¹ and
Miriam Vock²

¹Department of Education, University of Greifswald, Greifswald, Germany, ²Department of Education, University of Potsdam, Potsdam, Germany

Introduction: This study addresses the controversy over whether students consider grades accurate indicators of achievement. We focused on grades for oral participation, as teachers rely on informal assessments in assigning these grades, and educational measurement studies have questioned the validity of such assessments.

Methods: Based on two samples of university students (Sample 1: $M = 22$ years, 80% female; Sample 2: $M = 21$ years, 67% female) (Total $N = 431$), we measured whether students perceive grades for oral participation as being reliable indicators of their school achievement. We investigated variations in students' retrospective perceptions of oral participation grades in different school subjects. We also examined how students' retrospective perceptions of oral participation grades were linked to grading transparency and student achievement.

Results: Our findings indicate that students perceive oral participation grades as more accurate indicators of their achievement in languages than in mathematics. Oral participation grades were perceived as being more accurate indicators of student achievement by male students and students who reported greater transparency in the assignment of their grades. In mathematics, higher-achieving students perceived grades as being less valid indicators of their achievement.

Discussion: Our results imply that teachers should be mindful of transparency when assigning grades for oral participation. By increasing the transparency of their grading, for instance, by telling students in advance what aspects they factor into their grading, teachers can help students view grades for oral participation as valid indicators of their achievement and increase procedural and distributive justice.

KEYWORDS

teacher-assigned grades, grading practices, oral participation, informal assessment, transparency

1 Introduction

Teacher-assigned grades are ubiquitous in everyday school life, serving as the basis for important school career decisions, such as promotion to the next grade level (Westphal et al., 2020a), school track recommendation (Klapproth and Fischer, 2019), and university

admission (Zwick, 2019). In addition, grades function as feedback from the teacher to the student and can affect the student's academic self-concept, which is relevant to their further learning behavior (Marsh et al., 2017). When teachers assign grades, they are primarily required to assess the performance of their students. However, teachers also take other factors into account when assigning grades, such as the student's conscientiousness (Westphal et al., 2020b, 2021), academic motivation, and classroom behavior (Close, 2009; Hochweber et al., 2014; Westphal et al., 2016). Report card grades are therefore typically made up of assessments of more clearly structured situations, such as written performance on tests and project or presentation assessments (Organisation for Economic Co-operation and Development [OECD], 2013). However, various assessments of more informal performance situations, such as contributions to class discussions and the quality and quantity of general class participation, are also frequently included (Organisation for Economic Co-operation and Development [OECD], 2013).

Grades for informal performance situations are sometimes summarized as “grades for oral participation,” which can be understood as an umbrella term for various forms of active, mostly spontaneous, participation in class discussion, oral exams, or school project presentations (Krüll, 2023, p. 162–163). In the German school system, in which this study was conducted, grades for oral participation are supposed to account for around 50% of the report card grade, depending on the year level and domain (Falkenberg, 2020). These grades therefore contribute significantly to a student's overall grade and school career; nevertheless, there are few rules or specifications governing how grades for oral participation are determined (Falkenberg, 2020). Research has shown that deviations between teacher-assigned grades and standardized indicators of achievement are particularly pronounced when teachers place more emphasis on informal assessments (Martínez et al., 2009), including class participation. The fact that grades do not exclusively reflect student achievement has sparked a debate over whether they should play a significant role in centralized school career decisions and university admissions (Buckley et al., 2018; Hübner et al., 2020; Trautwein and Baeriswyl, 2007). Some also argue against grades being used as the sole basis for these decisions, as is the case with university placement in some countries, such as Germany. Less focus has been placed on the student perspective, and thus, on the question of how students perceive their grades.

In the present paper, we assessed the extent to which students perceive grades for oral participation as being an accurate indicator of their school achievement and how transparent they perceive these grades to be. We also examined differences in these perceptions across various school subjects. Given that oral skills are particularly relevant in languages (Geva, 2006; Sandlund et al., 2016), but less necessary in mathematics (Pace and Ortiz, 2015), we investigated whether students perceive grades for oral participation as more accurately reflecting their achievement in languages compared to mathematics. In addition, we investigated what it takes for students to perceive grades for oral participation as accurate indicators of their achievement; in particular, the role that perceived transparency plays in this regard. We also tested the hypothesis of whether higher-achieving students regard grades for oral participation—which are based to a lesser extent on actual achievement, and thus, on meritocratic rules—as being less appropriate indicators of their achievement.

1.1 What's in a school grade?

The question “What's in a grade?” (Bowers, 2011) has occupied researchers and educational policy stakeholders for many decades. Teacher education textbooks train teachers to base their grades on the achievement students show in the classroom (Brookhart, 2004; Linn and Miller, 2005). Teachers report using many different indicators when grading student achievement (Brookhart, 1993; McMillan, 2001; Randall and Engelhard, 2009). In assigning grades, teachers not only consider whether students have achieved the learning objectives set in class but also take into account their work habits and degree of effort (Brookhart, 1993; McMillan, 2001; Randall and Engelhard, 2009). According to teachers, however, the achievement that students show in class is the most important aspect of assigning grades (McMillan, 2001; Randall and Engelhard, 2009). These results on the teachers' perspective on grading are supported by large-scale educational studies that have investigated the interplay between teacher-assigned grades and student characteristics; for example, whether students who show more effort in class are awarded higher grades than those who show less effort (Hochweber et al., 2014).

Teacher-assigned grades reflect students' personality, motivation, and behavior in class, as well as their demographic characteristics (e.g., Boman, 2023; Hochweber et al., 2014; Krejtz and Nežlek, 2016; Spengler et al., 2013; Westphal et al., 2016, 2020b). Nevertheless, studies have shown that students' achievement on standardized tests largely explains teacher-assigned grades (Bowers, 2011). However, the extent to which grades reflect performance on standardized tests varies from teacher to teacher and is highly dependent on the form of assessment used (Martínez et al., 2009).

Another important finding of previous empirical research is that teachers differ in how appropriately they can evaluate their students' performance (for meta-analytical evidence see Hoge and Coladarsi, 1989; Südkamp et al., 2012). Thus, the appropriateness of school grades differs across teachers (e.g., Hochweber et al., 2014; Westphal et al., 2020b). This can be partially explained by class composition. For example, due to the in-class frame of reference, students in low-performing classes receive better grades for achievement than students in average-performing classes (Dompnier et al., 2006; Marsh, 1987). This big-fish-little-pond effect (Marsh, 1987) explains why teachers may overestimate or underestimate the performance of students in their class. Teachers' professional knowledge and beliefs can also influence their assessment practices (Lam, 2019; Yan et al., 2021), the accuracy of their assessments (Herppich et al., 2010; McElvany et al., 2012), and the appropriateness of the grades they assign (Brunner et al., 2013). Teachers have divergent opinions on the appropriateness of various assessment practices (Martínez et al., 2009). As they have some leeway in assigning grades, they employ a wide variety of assessment practices and base their grade assignments on more- or less-structured written tests, assessment worksheets, or homework assignments (Martínez et al., 2009). When teachers rely more heavily on informal assessments, e.g., class participation, in assigning grades, the grades are more likely to deviate from standardized achievement indicators (Martínez et al., 2009). Such deviations have sparked vigorous debates about whether the allocation of university and apprenticeship placements,

or even assignments to school tracks in multi-tiered school systems, should be based on school grades at all (Buckley et al., 2018; Hübner et al., 2020; Trautwein and Baeriswyl, 2007). However, less attention has been given to the extent to which students perceive grades for oral participation and performance as accurately reflecting their achievement.

1.2 Grades for oral participation

Active oral participation in class is graded implicitly or explicitly in many OECD countries and, in some countries, it constitutes a relevant portion of the report card grade. In a review of student assessment practices in different OECD countries, teachers disclosed that their end-of-year grades were composed of grades assigned for different achievement indicators, such as homework completion, presentations, and class participation (Organisation for Economic Co-operation and Development [OECD], 2013). Active participation is graded for several reasons. First, the development of oral-communicative skills, such as the ability to explain, argue, or discuss, is an important teaching goal in all subjects. In addition, teaching thrives on the active participation of students. Teachers may also want to give students the opportunity to compensate for a poor grade on a written test by earning good grades for regular oral participation, which may better reflect their actual achievement.

Martínez et al. (2009) reported that teachers in the United States view class participation as a more important factor in evaluating students than the child's achievement compared to the rest of the class and local or state-wide standards. In line with these findings, Bainbridge Frymier and Houser (2016) pointed out that "oral participation is generally highly valued in American classrooms and is often thought to be a good indicator of students' engagement in learning" (p. 83). In the United Kingdom, in 1990, a framework for oral assessment across different school levels was introduced into the National Curriculum for English (Department of Education and Science [DES], 1990). This framework was aimed at improving oracy, i.e., students' speaking and listening skills (Thompson, 2006; see also Voice 21, 2022), but Thompson (2006) critically noted that teachers' oral assessment practices seem to be based more on students' oral participation than on the quality of their language. In Germany, grades for oral participation constitute a relevant portion of the report card grade and their weighting in final year grades is 40–60% for most subjects (Krüll, 2023). Some federal states in Germany allow their teachers to determine final grades in some subjects almost entirely based on oral participation, e.g., 100% in the case of history in North Rhine-Westphalia, Germany's most populous state (Langhammer, 1998). Unlike the grading of written tests and exams, however, teachers are given little guidance or direction on how to assign oral participation grades. The core curriculum for secondary education in North Rhine-Westphalia stipulates only that "grading must take into account the quality, quantity, and continuity of contributions" (Ministry for School and Education of the State of North Rhine-Westphalia, 2019, p. 39). Thus, teachers can decide how to define oral participation, what criteria to use in grading it, and how to record performance (Krüll, 2023).

In empirical research, few studies have explicitly focused on grades for oral participation. A recently published interview study

of $N = 232$ teachers from North Rhine-Westphalia showed that in grading oral participation, the teachers primarily focused on the students' thinking, continuous performance, and technical correctness, weighting these criteria more or less equally for all students (Krüll, 2023); nevertheless, 77% of the teachers said they occasionally made exceptions to the use of these criteria. For example, in the case of high-achieving but introverted students, the teachers gave more weight to the quality of oral contributions than to their frequency. A substantial number of teachers also reported that they gave less weight to the oral participation of introverted students in determining their overall grades.

These findings are in line with criteria reported as relevant by 50 teachers in another German study by Mangold (2016). These teachers partially agreed that grades for oral participation were objective, meaning that another teacher would assign the same oral grade during the same evaluation process (Mangold, 2016). Teachers also partially agreed that grades for oral participation were reliable (meaning that repeating the assessment would yield the same oral grade) and valid (meaning that the oral grade actually reflected oral participation). In addition, teachers expressed partial agreement with the assertion that their students concur with the grades assigned for oral participation. Despite this, another teacher stated: "It is impossible to objectively and reliably evaluate oral participation" (Schöneberg, 2015, p. 1). Hence, to what extent is the teacher's view that students agree with their oral grades, as reported by Mangold (2016), true?

1.3 Students' perception of grades

According to Resh and Sabbagh (2016) "a sizeable portion of students seem to experience injustice in reward distribution in schools, both in grade allocations and in teacher-student relations, suggesting that schools are a meaningful source of injustice experiences for students" (p. 360). Yet, few studies have considered students' perspectives on performance assessment, teacher judgments, or teacher-assigned grades. In a series of studies, Resh and colleagues examined whether high school students perceived their grades as just by forming the quotient of the actual grade and their "seen" grade (the grade thought they should have earned) (Jasso and Resh, 2002; Resh, 2010, 2013; Resh and Dalbert, 2007). The authors showed that most students felt "under-rewarded," i.e., they felt they deserved better grades than they received.

In addition, few existing scales capture how students perceive performance assessments, teacher judgments, or teacher-assigned grades (Dorman and Knightley, 2006; Gerritsen-van Leeuwenkamp et al., 2018; Ibarra-Sáiz et al., 2021). The Perceptions of Assessment Tasks Inventory (PATI) (Dorman and Knightley, 2006) has been used most frequently. It comprises subscales that measure congruency with planned learning (i.e., whether teachers' assessments are congruent with the learning goals set in class), authenticity (i.e., whether teachers' assessments include real-life situations), student consultation (i.e., whether students know in advance which form of assessment will be used), transparency (i.e., whether students know the requirements of different assessment methods), and diversity (i.e., whether all students have an equal chance of successfully completing the assessment). The subscales of the PATI and instruments developed

by Gerritsen-van Leeuwenkamp et al. (2018) and Ibarra-Sáiz et al. (2021) refer to the practice of performance assessment by a teacher in their aggregate. They do not focus on how students perceive informal assessment practices. Overall, research on how students perceive informal assessment or grades for oral participation is particularly scarce. As a case in point, a literature search conducted for the present manuscript yielded 120 hits on the Web of Science database, of which 67 could be classified as educational research.¹ However, none of these studies focused on students' perceptions of informal classroom assessments or grades for oral participation.

In addition, grades based on informal assessment or oral participation in class may deviate more from standardized achievement indicators (Martínez et al., 2009). As a result, meritocratic rules are less strictly applied, which students—especially those with higher overall achievement—may perceive as being less fair (Baniyadi et al., 2023). Findings from Israel suggest that male students perceive their grades as being lower than those given to female students (Resh, 2013). In this context, the present study aimed to provide a deeper understanding of how students perceive oral participation grades and the factors that contribute to their perceptions.

1.4 The present study

Although students often orient their learning activities to what they need to accomplish in a course (Biggs et al., 2023; Close, 2009), very few studies have considered students' perspectives on performance assessment, teacher judgments, or teacher-assigned grades. Therefore, the first aim of the present study was to analyze the extent to which students perceive grades for oral participation as being an accurate indicator of their achievement. Since it is particularly important in informal forms of assessment, such as grades for oral participation, that students be informed in advance which achievement-related behaviors will be assessed, we also investigated how students perceive the transparency of oral participation grades:

RQ1a: To which extent do students perceive grades for oral participation as being an accurate indicator of their achievement?

RQ1b: How do students perceive the transparency of oral participation grades?

The second aim of this study was to examine whether these perceptions differ between different school subjects:

RQ2: Do students' perceptions differ between different school subjects?

H2: We hypothesize that students perceive grades for oral participation as being a more accurate indicator of their achievement in languages than in mathematics.

Thirdly, we investigated the factors that may affect students' perceptions regarding the validity of oral participation grades

as indicators of their achievement. Specifically, we studied how perceived transparency and student achievement were associated with the extent to which students perceived grades for oral participation as accurate indicators of their achievement. We controlled for gender, which is associated with students' perception of grades (Resh, 2013; Vogt, 2017). It has been argued that students' perception of grades depends on both the quality and transparency of assessment practices (Annerstedt and Larsson, 2010) and students' achievement in class (Baniyadi et al., 2023). As such, higher-achieving students may value grades less if the assignment of grades is less dependent on overall achievement in class (Baniyadi et al., 2023).

RQ3a: Do students perceive grades as better indicator of their achievement when they perceive greater transparency in the assignment of oral participation grades?

H3a: We hypothesize that students perceive grades for oral participation as being more accurate indicators of their achievement if they see greater transparency in the assignment of grades for oral participation.

RQ3b: Do students perceive grades as better indicator of their achievement when they achieve lower overall grades?

H3b: We hypothesize that students perceive grades for oral participation as being more accurate indicators of their achievement if they achieve lower overall grades.

2 Materials and methods

2.1 Samples

2.1.1 Sample 1

Sample 1 comprised $N = 155$ university students enrolled in different universities in Germany who participated in an online survey during the summer term of 2020. Students were on average 22 years old and in their first year of a bachelor's degree program; 80% were female (Supplementary Table 1).

2.1.2 Sample 2

Sample 2 comprised $N = 276$ university students enrolled at a German university who participated in an online survey between 2020 and 2023. Students were on average 21 years old and in their first or second year of a bachelor's degree program; 67% were female (Supplementary Table 1).

2.2 Measures

2.2.1 Perception of grades for oral participation

Based on the literature, we developed nine items to assess students' perceptions of grades for oral participation. The exact wording of the items is presented in Table 1 (for the German original, see Supplementary Table 2). First- or second-year university students were instructed to retrospectively answer the

¹ The search was based on the search string ("fairness of grades" OR "perception* of grades" OR "transparency of grades" OR "fairness of teacher judgment*" OR "perception* of teacher judgment*" OR "transparency of teacher judgment*" OR "fairness of assessment" OR "perception* of assessment" OR "transparency of assessment") NOT "assessment centers," and was conducted in April 2023.

items for their last two school years and separately for three school subjects: mathematics, German, and English as a Foreign Language (EFL). The items were answered on a 6-point Likert scale (1 = *do not agree at all* to 6 = *totally agree*). Prior to presenting the items, students were asked to indicate whether informal assessment took place in mathematics, German, and EFL and were only presented with the items for subjects for which they indicated that grades for oral participation were assigned.

2.2.2 End-of-semester grades and grade point average

Students reported the end-of-semester grades they had received in mathematics, German, and EFL at school. In the upper secondary years of the German school system, these grades are assigned in a different metric, ranging from 1 (unsatisfactory achievement) to 15 (excellent achievement). In addition, students reported the grade point average (GPA) they had received in their A-levels; these grades ranged from 1 (excellent achievement) to 6 (unsatisfactory achievement). We reverse-coded the grades so that higher grades represented better achievement.

2.2.3 Measure for validation: transparency of grades based on written assessment tests

We used three items from an inventory developed by [Ditton and Merz \(2013\)](#) and adapted their wording to assess the transparency of written assessment in the three school subjects mathematics, German, and EFL (e.g., “When I took exams at school, I usually knew exactly what was going to come up”). Items were answered on a 4-point Likert scale (1 = *do not agree at all* to 4 = *fully agree*). Reliability was acceptable for all three subjects (Cronbach’s $\alpha \geq 0.75$).

2.3 Statistical analyses

2.3.1 Exploratory factor analysis based on data from sample 1

To investigate the factor structure of our nine items assessing students’ perception of grades for oral participation, we performed an EFA with oblique Promax rotation on the Sample 1 data. Three separate EFAs were performed, one for each school subject. The number of items was determined by running parallel analysis and taking into account the screeplot and eigenvalues.

2.3.2 Confirmatory factor analysis based on data from sample 2

To confirm the factor structure, we ran a CFA on the Sample 2 data. In the first step, three separate CFAs were performed for each of the school subjects. In the second step, we specified a CFA in which we incorporated all three school subjects into one CFA model. For all subsequent analyses, we used the latent factor values of the scales. We also probed for measurement invariance across school subjects. Model fit was evaluated based on the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI values

TABLE 1 Results of the exploratory factor analysis.

	Item	Mathematics			German			EFL		
		M (SD)	AI	TR	M (SD)	AI	TR	M (SD)	AI	TR
1	The frequency of oral participation is a good indicator of achievement.	3.00 (1.43)	0.77	−0.12	3.13 (1.38)	0.54	0.11	3.44 (1.47)	0.60	0.10
2	In my opinion, it is a good idea to have grades for oral participation in addition to grades based on written assessment tests.	4.32 (1.34)	0.75	0.06	4.48 (1.34)	0.75	0.04	4.61 (1.32)	0.83	0.06
3	Grades for oral participation should be a more important part of the overall assessment.	2.50 (1.22)	0.74	−0.03	2.69 (1.24)	0.71	−0.08	2.84 (1.32)	0.68	−0.21
4	Grades for oral participation usually reflect my achievement.	3.93 (1.20)	0.54	0.21	3.99 (1.45)	0.74	−0.06	3.92 (1.46)	0.53	0.21
5	In mathematics/English/German, I think grades for oral participation are too subjective. [recoded] ^a	3.50 (1.49)	0.42	0.24	3.50 (1.35)	0.28	0.48	3.84 (1.29)	0.10	0.49
6	Often it was not clear to me how teachers determined grades for oral participation. [recoded]	3.19 (1.39)	−0.04	0.64	3.41 (1.30)	−0.06	0.65	3.64 (1.49)	−0.03	0.63
7	It was always openly explained to me how my final grade was made up.	3.98 (1.34)	−0.15	0.73	3.87 (1.46)	−0.10	0.62	3.78 (1.37)	−0.09	0.75
8	In general, grades for oral participation were assigned objectively.	3.23 (1.29)	0.05	0.56	3.33 (1.31)	−0.05	0.86	3.49 (1.27)	−0.06	0.74
9	Altogether, I had the impression of receiving fair grades. ^a	4.16 (1.17)	0.11	0.81	4.17 (1.31)	0.40	0.60	4.10 (1.24)	0.15	0.78

AI, Factor loading on the accurate indicator of achievement factor; TR, Factor loading on the transparency of grades for oral participation factor. ^aItem excluded. Higher loadings are printed in bold.

greater than or equal to 0.95, SRMR values greater than 0.08, and RMSEA values smaller than 0.06 indicate good model fit (Hu and Bentler, 1999; Wang and Wang, 2012). Measurement invariance is given when the differences in RMSEA between models are smaller than 0.015 and the differences in CFI values are smaller than 0.01.

2.3.3 One-way repeated measures ANOVA based on data from sample 2

To examine whether students' perceptions of grades for oral participation as an achievement indicator differed between school subjects, we computed a one-way repeated measures ANOVA with school subject as the within-subject factor. Pairwise paired *t*-tests between the different school subjects were computed and the *p*-values were adjusted using the Bonferroni correction method.

2.3.4 Regression analyses based on data from sample 2

We specified regression models in which student perception that grades for oral participation are an accurate indicator of achievement was regressed on student perception of the transparency of grades for oral participation, and actual teacher-assigned grades, controlling for student gender.

2.3.5 Missing data and software

Of our data, 19.2% of Sample 1 and 23.2% of Sample 2 were missing by design, because students had indicated that no grades for oral participation had been assigned. Beyond these missing values by design, 12% of the Sample 1 data and 0.8% of the Sample 2 data were missing. We used the full information maximum likelihood (FIML; Enders, 2001) approach when specifying the CFA models and conducting the regression analyses. The data were preprocessed using R and the CFA and regression analyses were conducted using the R package MplusAutomation (Hallquist and Wiley, 2018). The one-way repeated measures ANOVA was computed using the R package rstatix (Kassambara, 2023).

3 Results

3.1 Developing a measure for students' perception of teacher-assigned grades

In the first step, we conducted an EFA on the nine items based on the Sample 1 data for each school subject. Following O'Connor (2000), we used various indicators to determine how many factors should be retained. Parallel analysis indicated a two-factor solution for perceptions of grades in mathematics and German, and a three-factor solution for perceptions of grades in EFL (Figures 1A–C). For EFL, the eigenvalues also suggested a two-factor solution. We therefore extracted two factors for each subject. When extracting the two factors and comparing the solutions across school subjects, we found that two of our items showed loadings on different factors or high cross-loadings in at least one EFA solution (Table 1). Based on this analysis, we deleted the two items.

To investigate whether the two-factor solution would hold in a CFA, we computed a CFA on the seven items based on the Sample 2 data. After deleting one of the items, we found that the model fit the data well (Table 2). In the third step, we conducted a combined CFA specifying two factors for each of the three school subjects. This combined model also showed a good model fit (Table 3, configural model). We probed for measurement invariance across the three school subjects and found metric invariance (Table 3).

The first factor can be termed “achievement indicator,” reflecting the perception that grades for oral participation are a reliable indicator of achievement. The second factor can be termed “transparency,” reflecting the perception that grades for oral participation were assigned transparently. The bivariate correlations between the two latent factors were $r = 0.35$ in EFL and $r = 0.38$ in both mathematics and German, indicating that both factors were related to some extent (Table 4). The correlations for the latent factor “achievement indicator” across subjects were substantial ($0.52 \leq r \leq 0.68$), as was the correlation for the latent factor “transparency” across subjects ($0.45 \leq r \leq 0.56$) (Table 4).

The means, standard deviations, and reliabilities for the two scales are shown in Table 4. McDonald's Omega was acceptable for the achievement indicator construct and for the transparency construct in German and EFL, but was lower for the transparency construct in mathematics. Correlations between the latent factors and teacher-assigned grades showed that only the transparency construct was associated with grades (Table 5). Students with better grades in the respective school subject perceived the grades for oral participation as being more transparent. Correlations between the latent factors and transparency of formal assessment are also presented in Table 5. The transparency construct was associated with transparency of formal assessment in the respective subject. Students who reported greater transparency of formal assessment also perceived the grades for oral participation as being more transparent. However, correlations were in the medium range, indicating that perceived transparency in grades for oral participation and transparency in grading based on formal assessment depict different constructs.²

3.2 Differences in students' perceptions of grades for oral participation between subjects

To explore variations in student perceptions of the accuracy of oral participation grades as indicators of achievement between mathematics, EFL, and German, we ran a one-way repeated measures ANOVA with school subject as the within-subject factor (mathematics vs. German vs. EFL). The achievement indicator construct was statistically different for the different school subjects, $F(1, 340) = 103.76$, $p < 0.001$, generalized $\eta^2 = 0.04$. *Post-hoc* tests with Bonferroni adjustment indicated that all the pairwise differences showed statistically significant differences ($p < 0.001$). Thus, students perceived grades for oral participation as being a better indicator of their achievement in EFL than in German and as a better indicator of their achievement in German than in mathematics.

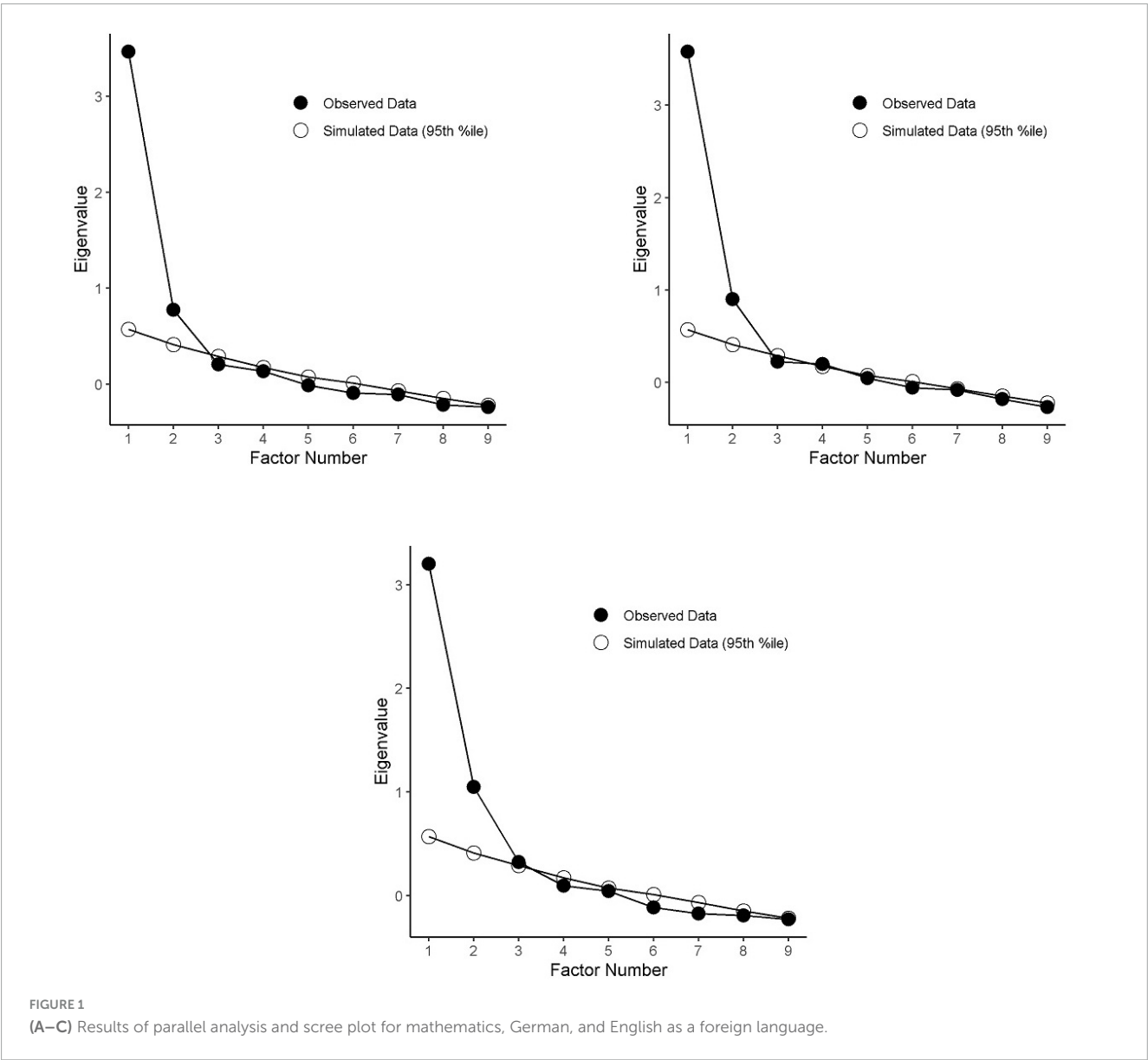


TABLE 2 Model fit of the confirmatory factor analysis for each school subject and intercorrelations.

	Free parameters	Model fit indices					
		AIC	BIC	CFI	TLI	RMSEA	SRMR
Mathematics	19	3555.278	3615.517	1.000	1.000	0.000	0.032
German	19	4495.614	4560.688	0.993	0.987	0.029	0.028
English	19	4523.524	4588.765	1.000	1.000	0.000	0.018

TABLE 3 Measurement invariance across school subjects.

	Free parameters	Model fit indices					
		AIC	BIC	CFI	TLI	RMSEA	SRMR
Configural	87	11900.726	12207.787	0.986	0.979	0.027	0.045
Metric	79	11891.111	12169.936	0.987	0.983	0.024	0.046
Scalar	70	11909.943	12157.003	0.968	0.958	0.038	0.054

TABLE 4 Mean, standard deviation, McDonald’s Omega, and intercorrelation of latent factors.

	<i>M</i>	<i>SD</i>	ω	Mathematics		German		EFL	
				MA AI	MA TR	GE AI	GE TR	EFL AI	EFL TR
MA AI	3.22	0.72	0.74	–					
MA TR	3.27	0.77	0.64	0.38***	–				
GE AI	3.36	0.79	0.74	0.68***	0.25**	–			
GE TR	3.50	0.79	0.68	0.26**	0.56***	0.38***	–		
EFL AI	3.59	0.75	0.73	0.52***	0.23*	0.57***	0.27**	–	
EFL TR	3.74	0.75	0.69	0.13	0.52***	0.10	0.45***	0.35***	–

MA, Mathematics; GE, German; AI, Grades for oral participation as an accurate indicator of achievement; TR, Transparency of grades for oral participation. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

TABLE 5 Intercorrelation of latent factors with grades and transparency of grades based on written assessment tests.

	Grades				Transparency of grades based on written assessment tests		
	GPA	Math	German	EFL	Mathematics	German	EFL
MA AI	–0.03	0.02	–0.07	–0.07	0.13	0.06	–0.07
MA TR	0.24***	0.26***	0.08	0.14*	0.39***	0.14*	0.21**
GE AI	–0.04	–0.02	–0.03	–0.06	0.04	0.13*	–0.12
GE TR	0.21***	0.16*	0.17**	0.18**	0.16*	0.38***	0.14*
EFL AI	0.01	0.04	0.01	0.11	0.13	0.10	0.15*
EFL TR	0.22***	0.19**	0.12	0.30***	0.32***	0.19**	0.46***

MA, Mathematics; GE, German; AI, Grades for oral participation as an accurate indicator of achievement; TR, Transparency of grades for oral participation. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

3.3 Factors affecting students’ perceptions of whether grades for oral participation are an accurate indicator of achievement

To examine the relationship between students’ achievement indicator perceptions, transparency perceptions, teacher-assigned grades, and gender, we computed regression models separately for each of the school subjects (Table 6). When taking into account student grades and gender, student transparency perceptions were a strong predictor of achievement indicator perceptions. Beyond transparency perceptions, male students perceived grades for oral participation as being a better indicator of achievement. In mathematics, students with lower end-of-semester grades perceived grades for oral participation as being a better indicator of achievement. In EFL and German, teacher-assigned grades were not significantly associated with the achievement indicator construct. The model explained 30% of the variance in students’ achievement indicator perceptions in mathematics and German and 31% of the variance in EFL.

4 Discussion

Students receive grades almost daily in school. Teachers assign grades based on a wide range of practices, from standardized tests to informal assessments (Martínez et al., 2009), and the evaluation of active and oral participation in class plays an important role in the process of grading. In Germany, for example, teachers are required to assess only around half of their report card grades

based on written performance (e.g., tests) and the other half based on oral participation. In addition, informal assessments play an important role in progressive teaching formats, which focus more on self-regulated learning and in which students pursue different learning objectives (Sliwka and Klopsch, 2022). Generally, grades are associated with numerous consequences for students, some of which are highly significant (e.g., Westphal et al., 2020a). However, little is known about how students perceive the assignment of grades. In particular, perceptions of informal assessments, such as grades for oral participation, have rarely been studied. In the present study, we addressed this research gap by measuring how students perceive grades for oral participation. We also examined variations in these perceptions between different school subjects. Moreover, this study investigated the factors that contribute to the perception of grades for oral participation as being an accurate indicator of achievement.

4.1 Do students perceive grades for oral participation as being valid indicators of their achievement?

The students in our study only partially agreed that grades for oral participation are valid indicators of their achievement. However, the perceptions that oral participation grades are a more or less valid indicator of achievement varied for different school subjects. Students perceived grades for oral participation in EFL as being a more valid indicator of achievement, while they perceived them as being a less valid indicator of achievement in mathematics. This can be explained by the fact that learning foreign languages is

TABLE 6 Regressing students' perception of grades for oral participation as a valid indicator of achievement by students' perception of transparency, grade, and gender.

	<i>b</i>	<i>SE</i>	95% CI		β	<i>p</i>
			<i>LL</i>	<i>UL</i>		
Mathematics						
Intercept	1.86	0.18	1.50	2.22		0.000
Mathematics transparency	0.46	0.05	0.36	0.56	0.50	0.000
Mathematics grade	−0.02	0.01	−0.05	0.00	−0.12	0.037
Gender ¹	0.29	0.08	0.12	0.45	0.19	0.001
Adjusted <i>R</i> ²	29.9					
German						
Intercept	1.78	0.22	1.35	2.20		0.000
German transparency	0.48	0.05	0.38	0.59	0.50	0.000
German grade	−0.02	0.01	−0.05	0.01	−0.08	0.136
Gender ¹	0.31	0.09	0.13	0.48	0.19	0.001
Adjusted R2	30.4					
English as a Foreign Language						
Intercept	1.71	0.21	1.30	2.13		0.000
EFL transparency	0.50	0.05	0.39	0.61	0.52	0.000
EFL grade	−0.01	0.01	−0.03	0.02	−0.03	0.604
Gender ¹	0.29	0.08	0.13	0.46	0.19	0.001
Adjusted R2	30.8					

¹Reference: female.

particularly concerned with oral skills (Geva, 2006; Sandlund et al., 2016), whereas these are also necessary in mathematics (Pace and Ortiz, 2015), but to a much lesser extent.

4.2 What contributes to students' perceptions that grades for oral participation are a valid indicator of achievement?

When examining the factors that contribute to students' perception that grades for oral participation are a valid indicator of their achievement, our findings showed that transparency, i.e., the perception that grades for oral participation are assigned transparently, was particularly crucial. This finding is consistent with the notion that the perception of grades and their legitimacy largely depends on the quality and transparency of assessment practices (Annerstedt and Larsson, 2010). Whereas students' perception of grades as being a valid indicator of achievement has been framed as reflecting distributive justice (Resh and Sabbagh, 2017), students' perception of grading transparency could be seen as reflecting procedural justice. Procedural justice—defined as “evaluations regarding the justness of the processes (or means) by which these resources are distributed” (Resh and Sabbagh, 2017, p. 390)—is relevant for students' school engagement and community volunteering (see Resh and Sabbagh, 2017), although these authors operationalized it differently than we did in our study. In addition, greater distributive justice of grade allocation is linked

to lower levels of academic dishonesty and violence (Resh and Sabbagh, 2017). While our findings indicate that these two aspects of justice are closely intertwined, each may be of distinctive value for transmitting civic norms that are crucial in a democratic society (Resh and Sabbagh, 2017).

In mathematics, students with lower end-of-semester grades perceived grades for oral participation as being more valid indicators of their achievement. This finding may reflect the notion that higher-achieving students value grading practices that are closely aligned with meritocratic rules (Baniasadi et al., 2023). In mathematics, grades for oral participation may diverge to a greater extent from more standardized forms of classroom assessment, which could explain our finding that higher-achieving students consider grades for oral participation as being less valid. In contrast, lower-achieving students—i.e., those who score lower on more standardized and written test formats in mathematics—may value the chance to improve their grades in mathematics by contributing to class discussions or generally participating actively in class. In German and EFL, end-of-semester grades were not associated with students' perception that grades for oral participation are an accurate indicator of their achievement. It is possible that—in these subjects, at least—grades for oral participation that reflect the quality and quantity of general active participation in class are more closely aligned with scores based on standardized tests. Students who show better overall achievement in languages may therefore not consider grades for oral participation in these subjects as diverging from meritocratic principles.

Finally, female students perceived grades for oral participation as being less valid indicators of their achievement. This finding

contradicts previous research indicating that if gender differences in perception of grades exist, it is male students who feel that grades are more unjust (Correia and Dalbert, 2007; Jasso and Resh, 2002; Resh and Dalbert, 2007). Jasso and Resh (2002) interpreted these feelings as reflecting (somewhat inconsistent) findings indicating that female students receive more favorable grades (e.g., Heyder et al., 2017; but see also Westphal et al., 2016). These studies on male students' greater perceptions of injustice in their grades focused on overall grades and did not control for actual end-of-term grades, which may explain the differential findings of the present study. Our finding that female students perceive grades for oral participation as being a less valid indicator of their achievement than male students seems more straightforward for mathematics than for languages. In mathematics, female students have lower expectations for success (Parker et al., 2020) and may therefore feel less confident about active class participation even when they show similar levels of achievement as male students. In languages, however, female students have higher expectations for success than male students (Parker et al., 2020). In addition, previous findings indicate that extraversion—associated with greater class participation—is beneficial for female students' overall grades in languages, but not for male students (Spinath et al., 2010). Therefore, it is surprising that female students perceived grades for oral participation as being less valid indicators of their achievement not only in mathematics but also in languages. Qualitative interviews could be informative in explaining why female students perceive grades for oral participation as less adequately representing their achievement.

4.3 Limitations and implications for future research

Our study exhibits some limitations. First, we only assessed students' perceptions retrospectively, which may introduce a recall bias. Future studies should apply our instruments in an active school context, taking into account the shared perceptions of all students in the classroom, information about the grading practices of the teacher (e.g., frequency of giving grades for oral participation), and ideally, external ratings of the transparency of informal assessments. In addition, our findings are based on informal assessment practices that are common in the German school system. Informal assessment practices in primary school and other school systems may differ from those in higher secondary schools in Germany. Therefore, our findings are not easily generalizable to the assessment of younger students and other countries. Finally, our study does not consider the extent to which aspects of the teacher–student relationship or perceived instructional quality affect perceptions of informal assessments.

5 Conclusion and educational implications

The validity of grades has often been studied from an educational measurement perspective (Martínez et al., 2009). Based on this research, the incremental validity of grades above and beyond standardized test scoring has been a matter of controversy

(Zwick, 2019). The present study sheds light on the student's perspective on this issue. This study focused on perceptions of grades for oral participation, where teachers rely more heavily on informal assessments, the validity of which has been particularly questioned in educational measurement studies. We found that in mathematics in particular, higher-achieving students perceived grades for oral participation as being less valid indicators of their achievement. In languages, students—regardless of their level of achievement—perceived grades for oral participation as being more accurate indicators of their achievement. The transparency with which grades for oral participation are assigned is closely related to students' perceptions of how valid they are. Our findings have important implications for questions on the distributive justice of school grades.

The predictive power of high school grades for college success is higher than that of admission tests, in part because grades do not exclusively reflect achievement (Galla et al., 2019). At the same time, this aspect raises the question of distributive justice when important decisions, such as admission to university, are based on grades assigned by teachers. To increase procedural and distributive justice, instructors should be mindful of transparency, especially when using informal assessments to assign grades, such as when assigning grades for oral participation. By increasing the transparency of their grading, for instance, by telling students in advance what aspects they factor into their grading, teachers can help students view grades for oral participation as valid indicators of their achievement. Another aspect that needs to be considered is that performance grading can reduce intrinsic motivation, as it can undermine the basic needs for autonomy, competence, and connectedness (Krijgsman et al., 2017), which, according to self-determination theory, are fundamental for intrinsic motivation (Deci and Ryan, 2000). Yet, “the motivational impact of grading is likely to depend on its functional significance” (Krijgsman et al., 2017, p. 202; Vansteenkiste et al., 2008). If students perceive grades less as information about their learning and more as a performance evaluation, this may have greater affective and motivational consequences (e.g., Ryan and Brown, 2005). As such, students may feel pressured to perform well (autonomy frustration), might feel like a failure when receiving or anticipating low grades (competence frustration), or may feel rejected when anticipating their peers' reactions to a bad grade (relatedness frustration; Krijgsman et al., 2017). Accordingly, it may be particularly important that teachers “induce feelings of choice or freedom, feelings of connection to others, and opportunities to reach criteria (i.e., need satisfaction), while also reducing feelings of pressure to perform well, feelings of rejection by others, and feelings of failure (i.e., need frustration) in order to stimulate positive motivational and affective experiences” (Krijgsman et al., 2017). If the legal regulations in the school system and the circumstances at the individual school allow it, it may be in the best interest of learners to use alternative forms of performance feedback, such as competency grids, grading rubrics, narrative assessments, and student self-assessment techniques. These approaches could better meet basic needs according to self-determination theory and could also be beneficial in the context of civic education. If performance grading is indispensable because grades are needed, e.g., for selection and allocation decisions (Westphal, 2024), it is essential that grades are assigned transparently. This is crucial because the perception by students that grades are assigned in a fair and just manner

can be conducive to students' civic behavior (Resh and Sabbagh, 2017). Most of all, however, the transparency of the assignment of grades—in terms of perceived distributive justice—has a value of its own.

Data availability statement

The data used for the current study are available from the corresponding author upon reasonable request.

Ethics statement

Ethical approval was not required for the studies involving humans because the Ethics Committee of the University of Potsdam issues ethics votes for planned intervention studies or studies in which vulnerable persons are examined or in which the participants are not informed (objectives of the study must be concealed) or in which there is pressure to commit, so that it cannot be ensured that participants participate voluntarily. None of these points were present in this case. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AW: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. CH: Formal Analysis, Methodology, Writing – review & editing. MV: Conceptualization, Investigation, Resources, Writing – review & editing.

References

- Annerstedt, C., and Larsson, S. (2010). 'I have my own picture of what the demands are': Grading in Swedish PEH – problems of validity, comparability and fairness. *Eur. Phys. Educ. Rev.* 16, 97–115. doi: 10.1177/1356336X10381299
- Bainbridge Frymire, A., and Houser, M. L. (2016). The role of oral participation in student engagement. *Commun. Educ.* 65, 83–104. doi: 10.1080/03634523.2015.1066019
- Baniasadi, A., Salehi, K., Khodaie, E., Bagheri Noaparast, K., and Izanloo, B. (2023). Fairness in classroom assessment: A systematic review. *Asia Pac. Educ. Res.* 32, 91–109. doi: 10.1007/s40299-021-00636-z
- Biggs, J., Tang, C., and Kennedy, G. (2023). *Teaching for Quality Learning at University*, 5th Edn. New York, NY: McGraw-Hill Education.
- Boman, B. (2023). The influence of SES, cognitive, and non-cognitive abilities on grades: Cross-sectional and longitudinal evidence from two Swedish cohorts. *Eur. J. Psychol. Educ.* 38, 587–603. doi: 10.1007/s10212-022-00626-9
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educ. Res. Eval.* 17, 141–159. doi: 10.1080/13803611.2011.597112
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *J. of Educ. Meas.* 30, 123–142. doi: 10.1111/j.1745-3984.1993.tb01070.x
- Brookhart, S. M. (2004). *Grading*. London: Pearson.
- Brunner, M., Anders, Y., Hachfeld, A., and Krauss, S. (2013). "The diagnostic skills of mathematics teachers," in *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*, eds M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (Berlin: Springer), 229–248.
- Buckley, J., Letukas, L., and Wildavsky, B. (eds) (2018). *Measuring Success: Testing, Grades, and the Future of College Admissions*. Baltimore, MA: John Hopkins University Press.
- Close, D. (2009). Fair grades. *Teach. Philos.* 32, 361–398. doi: 10.5840/teachphil200932439
- Correia, I., and Dalbert, C. (2007). Belief in a just world, justice concerns, and well-being at Portuguese schools. *Eur. J. Psychol. Educ.* 22, 421–437. doi: 10.1007/BF03173464
- Dalbert, C. (2000). "Gerechtigkeitskognitionen in der Schule [Perceptions of justice in school]," in *Handlungsleitende Kognitionen in der pädagogischen Praxis [Cognitions that Guide Action in Educational Practice]*, eds C. Dalbert and E. J. Brunner (Baltmannsweiler: Schneider Verlag Hohengehren), 3–12.
- Deci, E. L., and Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inquiry* 11, 227–268. doi: 10.1207/S15327965PLI1104_01
- Department of Education and Science [DES]. (1990). *English in the National Curriculum (No. 2)*. Dublin: HMSO.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1522695/full#supplementary-material>

- Ditton, H., and Merz, D. (2013). *QuASSU: QualitätsSicherung in Schule und Unterricht - Fragebogenerhebung Erhebungszeitpunkt 2 [Quality Assurance in Schools and Teaching - Questionnaire Survey Survey Date 2]*. Germany: Forschungsdatenzentrum Bildung am DIPF. doi: 10.7477/18.31:1
- Dompnier, B., Pansu, P., and Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *Eur. J. Psychol. Educ.* 21, 119–133. doi: 10.1007/BF03173572
- Dorman, J. P., and Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educ. Stud.* 32, 47–58. doi: 10.1080/03055690500415951
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychol. Methods* 6, 352–370. doi: 10.1037/1082-989X.6.4.352
- Falkenberg, K. (2020). *Gerechtigkeitsüberzeugungen bei der Leistungsbeurteilung: Eine Grounded-Theory-Studie mit Lehrkräften im Deutsch-schwedischen Vergleich [Fairness Beliefs in Performance appraisal: A Grounded Theory Study with Teachers in a German-Swedish Comparison]*. Berlin: Springer VS.
- Galla, B. M., Shulman, E. P., Plummer, B. D., Gardner, M., Hutt, S. J., Goyer, J. P., et al. (2019). Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. *Am. Educ. Res. J.* 56, 2077–2115. doi: 10.3102/0002831219843292
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., and Kester, L. (2018). Developing questionnaires to measure students' expectations and perceptions of assessment quality. *Cogent Educ.* 5:1464425. doi: 10.1080/2331186X.2018.1464425
- Geva, E. (2006). "Second-language oral proficiency and second-language literacy," in *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, eds D. August and T. Shanahan (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 123–139.
- Hallquist, M. N., and Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Struct. Equ. Model. Multidiscipl. J.* 25, 621–638. doi: 10.1080/10705511.2017.1402334
- Herppich, S., Wittwer, J., Nückles, M., and Renkl, A. (2010). Do tutors' content knowledge and beliefs about learning influence their assessment of tutees' understanding? *Proc. Annu. Meet. Cogn. Sci. Soc.* 32, 314–319.
- Heyder, A., Kessels, U., and Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions? *Br. J. Educ. Psychol.* 87, 205–223. doi: 10.1111/bjep.12145
- Hochweber, J., Hosenfeld, I., and Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *J. Educ. Psychol.* 106, 289–300. doi: 10.1037/a0033829
- Hoge, R. D., and Coladarc, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Rev. Educ. Res.* 59, 297–313. doi: 10.3102/00346543059003297
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscipl. J.* 6, 1–55. doi: 10.1080/10705519909540118
- Hübner, N., Wagner, W., Hochweber, J., Neumann, M., and Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *J. Educ. Psychol.* 112, 204–220. doi: 10.1037/edu0000351
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., and Boud, D. (2021). The quality of assessment tasks as a determinant of learning. *Assess. Eval. High. Educ.* 46, 943–955. doi: 10.1080/02602938.2020.1828268
- Jasso, G., and Resh, N. (2002). Exploring the sense of justice about grades. *Eur. Sociol. Rev.* 18, 333–351. doi: 10.1093/esr/18.3.333
- Kassambara, A. (2023). *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests (Version 0.7.2) [Computer software]*. Available online at: <https://CRAN.R-project.org/package=rstatix>.
- Klapproth, F., and Fischer, B. D. (2019). Preservice teachers' evaluations of students' achievement development in the context of school-track recommendations. *Eur. J. Psychol. Educ.* 34, 825–846. doi: 10.1007/s10212-018-0405-x
- Krejtz, I., and Nezek, J. B. (2016). It's Greek to me: Domain specific relationships between intellectual helplessness and academic performance. *J. Soc. Psychol.* 156, 664–668. doi: 10.1080/00224545.2016.1152219
- Krijgsman, C., Vansteenkiste, M., van Tartwijk, J., Maes, J., Borghouts, L., Cardon, G., et al. (2017). Performance grading and motivational functioning and fear in physical education: A self-determination theory perspective. *Learn. Individ. Diff.* 55, 202–211. doi: 10.1016/j.lindif.2017.03.017
- Krüll, C. (2023). *Wie Erfassen und Bewerten Lehrkräfte Mündliche Mitarbeit? [How do Teachers Assess and Grade Oral Participation?]*. Hildesheim: Georg Olms Verlag.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System* 81, 78–89. doi: 10.1016/j.system.2019.01.006
- Langhammer, R. (1998). Mündliche Noten“ im Geschichts- und Politikunterricht: Erhebungs- und Bewertungsvorschläge [“Oral grades” in history and politics lessons: survey and assessment suggestions]. *German J. History Politics Their Didactics* 26, 38–50.
- Linn, R. L., and Miller, M. D. (2005). *Measurement and Assessment in Teaching*, 9th Edn. Hoboken, NJ: Pearson Prentice Hall.
- Mangold, A. (2016). Die Bewertung Mündlicher Mitarbeit: Eine Empirische Studie zu Anspruch und Wirklichkeit der schulischen Bewertungspraxis [The Assessment of Oral Participation: An Empirical Study on the Claim and Reality of School Assessment Practices]. Kassel: Zentrum für Lehrerbildung der Universität Kassel.
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *J. Educ. Psychol.* 79, 280–295. doi: 10.1037/0022-0663.79.3.280
- Marsh, H. W., Martin, A. J., Yeung, A. S., and Craven, R. G. (2017). "Competence self-perceptions," in *Handbook of Competence and Motivation: Theory and Application*, eds A. J. Elliot, C. S. Dweck, and D. S. Yeager (New York, NY: The Guilford Press), 85–115.
- Martínez, J. F., Stecher, B., and Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educ. Assess.* 14, 78–102. doi: 10.1080/10627190903039429
- McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H., and Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: Teachers' diagnostic skills, pedagogical beliefs and motivation. *Eur. J. Psychol. Educ.* 27, 403–420. doi: 10.1007/s10212-011-0078-1
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educ. Meas. Issues Pract.* 20, 20–32. doi: 10.1111/j.1745-3992.2001.tb00055.x
- Ministry for School and Education of the State of North Rhine-Westphalia (2019). *Deutsch: Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen [German: Core Curriculum for Lower Secondary School in North Rhine-Westphalia]*. Available online at: https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/196/g9_d_klp_%203409_2019_06_23.pdf (accessed October 4, 2024).
- O'Connor, B. P. (2000). Spss and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instruments Comput.* 32, 396–402. doi: 10.3758/bf03200807
- Organisation for Economic Co-operation and Development [OECD]. (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. Paris: OECD, doi: 10.1787/9789264190658-en
- Pace, M. H., and Ortiz, E. (2015). Oral language needs: Making math meaningful. *Teach. Children Mathem.* 21, 494–500. doi: 10.5951/teachchildmath.21.8.0494
- Parker, P. D., van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., and Noetel, M. (2020). The intersection of gender, social class, and cultural context: A meta-analysis. *Educ. Psychol. Rev.* 32, 197–228. doi: 10.1007/s10648-019-09493-1
- Randall, J., and Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *J. Educ. Res.* 102, 175–186. doi: 10.3200/JOER.102.3.175-186
- Resh, N. (2010). Sense of justice about grades in school: Is it stratified like academic achievement? *Soc. Psychol. Educ.* 13, 313–329. doi: 10.1007/s11218-010-9117-z
- Resh, N. (2013). "Grades and sense of justice about grades: The gender aspect," in *Academic Achievement: Predictors, Learning Strategies and Influences of Gender*, eds L. Zhang and J. Chen (Hauppauge, NY: Nova Science Publishers), 95–112.
- Resh, N., and Dalbert, C. (2007). Gender differences in sense of justice about grades: A comparative study of high school students in Israel and Germany. *Teach. Coll. Rec.* 109, 322–342. doi: 10.1177/016146810710900206
- Resh, N., and Sabbagh, C. (2016). "Justice and education," in *Handbook of Social Justice Theory and Research*, eds C. Sabbagh and M. Schmitt (Berlin: Springer), 349–367.
- Resh, N., and Sabbagh, C. (2017). Sense of justice in school and civic behavior. *Soc. Psychol. Educ.* 20, 387–409. doi: 10.1007/s11218-017-9375-0
- Ryan, R. M., and Brown, K. W. (2005). "Legislating competence: High-stakes testing policies and their relations with psychological theories and research," in *Handbook of Competence and Motivation*, eds A. J. Elliot and C. S. Dweck (New York, NY: The Guilford Press), 354–372.
- Sandlund, E., Sundqvist, P., and Nyroos, L. (2016). Testing L2 talk: A review of empirical studies on second-language oral proficiency testing. *Lang. Linguistics Compass* 10, 14–29. doi: 10.1111/lnc3.12174
- Schöneberg, D. (2015). *Meine Noten sind ungerecht [My grades are not fair]. Bildungslücken [Education gaps]*. Available online at: <https://bildungsluecken.net/210-meine-noten-sind-ungerecht> (accessed October 4, 2024).
- Sliwka, A., and Klopsch, B. (2022). *Deeper Learning in der Schule: Pädagogik des digitalen Zeitalters [Deeper Learning in School: Pedagogy in the digital age]*. Weinheim: Beltz.
- Spengler, M., Lüdtke, O., Martin, R., and Brunner, M. (2013). Personality is related to educational outcomes in late adolescence: Evidence from two large-scale achievement studies. *J. Res. Pers.* 47, 613–625. doi: 10.1016/j.jrp.2013.05.008
- Spinath, B., Freudenthaler, H. H., and Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Pers. Individ. Diff.* 48, 481–486. doi: 10.1016/j.paid.2009.11.028

- Südkamp, A., Kaiser, J., and Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *J. Educ. Psychol.* 104, 743–762. doi: 10.1037/a0027627
- Thompson, P. (2006). Towards a sociocognitive model of progression in spoken English. *Cambridge J. Educ.* 36, 207–220. doi: 10.1080/03057640600718596
- Trautwein, U., and Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen [When high-achieving classmates are a disadvantage: Reference group effects in transfer decisions]. *German J. Educ. Psychol.* 21, 119–133. doi: 10.1024/1010-0652.21.2.119
- Vansteenkiste, M., Ryan, R. M., and Deci, E. L. (2008). "Self-determination theory and the explanatory role of psychological needs in human well-being," in *Capabilities and Happiness*, eds L. Bruni, F. Comim, and M. Pugno (Oxford: Oxford University Press), 187–223. doi: 10.1017/S0266267111000058
- Vogt, B. (2017). *Just Assessment in School: A context-Sensitive Comparative Study of Pupils' Conceptions in Sweden and Germany* (Publication No. 298/2017). [Doctoral dissertation]. Växjö: Linnaeus University Press.
- Voice 21. (2022). We're on a Mission to Transform the Learning and Life Chances of Young People Through Talk. Voice 21. Available online at: <https://voice21.org/>
- Wang, J., and Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. Hoboken, NJ: John Wiley & Sons.
- Westphal, A. (2024). Pädagogische Professionalität von Lehrkräften beim Beurteilen von Schülerleistungen: Ein Forschungsüberblick aus kompetenztheoretischer Perspektive [Pedagogical expertise of teachers in assessing student achievement: a research overview from a competence-oriented perspective]. *Pädagogische Rundschau*, 78, 579–591. doi: 10.3726/PR052024.0047
- Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., and McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemp. Educ. Psychol.* 46, 218–227. doi: 10.1016/j.cedpsych.2016.06.004
- Westphal, A., Lazarides, R., and Vock, M. (2020b). Are some students graded more adequately than others? Student characteristics as moderators of the relationships between teacher-assigned grades and test scores. *Br. J. Educ. Psychol.* doi: 10.1111/bjep.12397
- Westphal, A., Vock, M., and Kretschmann, J. (2021). Unravelling the relationship between teacher-assigned grades, student personality, and standardized test scores. *Front. Psychol.* 12:627440. doi: 10.3389/fpsyg.2021.627440
- Westphal, A., Vock, M., and Lazarides, R. (2020a). Are more conscientious seventh- and ninth-graders less likely to be retained? Effects of Big Five personality traits on grade retention in two different age cohorts. *J. Appl. Dev. Psychol.* 66:101088. doi: 10.1016/j.appdev.2019.101088
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., and Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assess. Educ. Principles Policy Pract.* 28, 228–260. doi: 10.1080/0969594X.2021.1884042
- Zwick, R. (2019). Assessment in American higher education: The role of admissions tests. *Ann. Am. Acad. Polit. Soc. Sci.* 683, 130–148. doi: 10.1177/0002716219843469