



When teachers' assessment intentions meet students' classroom experience: Evidence of interpretive misalignment in secondary physical education

Ernesto Panadero^{a,b,c,1} , Vasiliki Pitsia^{d,2}, Javier Fernández-Ruiz^{e,3}, Daniel García-Pérez^{f,*,4}, Rodrigo Pardo^{g,5} 

^a Centre for Assessment Research, Policy and Practice in Education (CARPE), School of Policy and Practice, Institute of Education, St. Patrick's Campus, Dublin City University, Dublin, Ireland

^b Education Regulated Learning and Assessment (ERLA group), Facultad de Educación y Deportes, Universidad de Deusto, Bilbao, Spain

^c IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

^d Educational Research Centre, Dublin, Ireland

^e Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, Spain

^f Departamento de Investigación y Psicología en Educación, Facultad de Educación-CFP, Universidad Complutense de Madrid, Spain

^g Facultad de Ciencias de la Actividad Física y del Deporte (INEF), Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Keywords:

Assessment conceptions
Classroom climate
Multilevel modelling
Student perceptions
Physical education

ABSTRACT

Assessment operates in classrooms not only as an instructional procedure but also as a communicative process through which teachers convey meanings about learning and performance. Yet little is known about whether teachers' conceptions of assessment are reflected in their students' perceptions of assessment and classroom motivational climate. This study examined this cross-level relationship by linking teacher conceptions and characteristics to students' perceptions of assessment and classroom motivational climate in secondary physical education. Data were collected from 989 students taught by 17 teachers in Spanish secondary schools and were analysed through a series of two-level linear regression models. A significant proportion of variance in each student-reported outcome was attributed to between-teacher differences. Teacher age and self-reported formative orientation were the most consistent predictors. Notably, stronger self-reported formative orientation by teachers was associated with less positive student perceptions on some assessment dimensions. These findings suggest that the meanings of assessment are not necessarily shared in the same way by teachers and students, highlighting assessment as a relational and interpretive process in which students construct the meaning of evaluation through classroom interaction.

Classroom assessment plays a central role in shaping how students engage with learning, interpret their progress, and understand their relationship with their teachers (Brown, 2017). Although often treated as a technical component of instruction, assessment also functions as a communicative practice. Through feedback, grading, and evaluation routines, teachers convey implicit messages about what counts as success, what is valued in learning, and what students are expected to do (Black & Wiliam, 2009; Brookhart, 2011). Consequently, assessment

does not only measure learning; it contributes to how students make sense of their learning environment and their place within it.

Teachers' conceptions of assessment, understood as their underlying beliefs about the purposes and functions of assessment (Brown, 2008), influence how assessment is enacted in classrooms. Research on conceptions of assessment has shown that teachers differ in how they understand the purpose of evaluating students. Some emphasize assessment as a means of supporting learning and improvement,

* Correspondence to: Departamento de Investigación y Psicología en Educación, Universidad Complutense de Madrid, Facultad de Educación-CFP, Calle Rector Royo-Villanova, 1, Madrid 28040, Spain.

E-mail address: daniel.garcia.perez@ucm.es (D. García-Pérez).

¹ <https://orcid.org/0000-0003-0859-3616>

² <https://orcid.org/0000-0002-8172-0397>

³ <https://orcid.org/0000-0001-5419-7687>

⁴ <https://orcid.org/0000-0002-5697-1792>

⁵ <https://orcid.org/0000-0002-6558-9915>

whereas others view it primarily as a mechanism for accountability or certification (Brown, 2008; Harris & Brown, 2009). These conceptions are assumed to guide classroom practices. However, students do not have direct access to teachers' beliefs; they experience assessment only through the interactions, routines, and expectations that structure classroom life. As a result, the meaning of assessment in a classroom may depend not only on what teachers intend to communicate but also on how students interpret those practices.

Despite extensive work on teacher conceptions and, separately, on students' perceptions of assessment, relatively little research has examined whether teachers' understandings of assessment are actually reflected in their students' perceptions of assessment and classroom motivational climate. This question is particularly relevant in secondary education, where assessment practices are more formalized and students are highly sensitive to evaluation and comparison (Harlen & Crick, 2003). If assessment operates as a communicative and relational process, then different classrooms might provide systematically different assessment experiences to students, even within the same curricular context.

The present study addresses this issue by adopting a cross-level perspective linking teachers' characteristics to students' perceptions of assessment and classroom motivational climate. Using matched data from secondary physical education teachers and their students, we examine whether students' perceptions vary between teachers and whether such differences are associated with teachers' assessment conceptions and demographic characteristics.

1. Teacher conceptions of assessment

Teachers' conceptions of assessment influence how they design, enact, and communicate evaluation practices in the classroom (Fernández-Ruiz & Panadero, 2020; Postareff et al., 2012). These conceptions refer to teachers' underlying beliefs about the purposes and consequences of assessing students (Brown, 2008). Some teachers emphasize assessment as a means of supporting learning and improvement, whereas others view it primarily as a mechanism for certification, grading, or accountability (Harris & Brown, 2009; Looney, 2011).

Importantly, teachers rarely hold a single, coherent orientation. Rather, their views tend to be hybrid and context-dependent, shaped by institutional demands, curriculum requirements, and personal pedagogical philosophies (Brown et al., 2024; DeLuca et al., 2025; Remesal, 2007). As a result, assessment intentions are negotiated within practical constraints, and teachers may simultaneously endorse formative principles while operating in systems that require summative judgments.

Brown's (2008) framework distinguishes several broad conceptions of assessment, including assessment for improvement and assessment for accountability. These orientations have implications for how teachers structure classroom activities, provide feedback, and involve students in evaluation processes. Teachers who emphasize improvement tend to value feedback, adaptation of instruction, and student participation, whereas accountability-oriented perspectives often prioritize grading, documentation, and certification (Harris & Brown, 2009; Wiliam, 2011).

However, a growing body of research suggests that teachers' stated beliefs do not always correspond closely to their enacted practices. Even teachers who report formative orientations may rely on grade-focused routines, limited feedback, or test-dominated evaluation structures (Bennett, 2011). This discrepancy is not necessarily a contradiction but reflects the influence of contextual pressures such as time constraints, class size, or institutional requirements (Panadero et al., 2014). Consequently, students do not encounter teachers' beliefs directly; they encounter patterns of interaction, feedback, and expectations that only partially reflect those beliefs.

From the student perspective, assessment is therefore experienced through classroom events rather than teacher intentions. If teachers' conceptions shape classroom routines, but those routines are interpreted

by students within ongoing interactions, the meaning of assessment in a classroom cannot be inferred solely from teachers' reported beliefs. Instead, it must be examined in relation to how students perceive and respond to assessment practices. This raises the question of whether differences in teachers' conceptions are actually reflected in students' perceptions.

2. Students' interpretation of classroom assessment and climate

Students do not experience assessment as a set of abstract principles but through everyday classroom events. Their perceptions are shaped by how assessment is implemented, how feedback is communicated, and how evaluation practices are embedded within classroom interactions (Brookhart & Durkin, 2003; Brown & Hirschfeld, 2008; Struyven et al., 2005). Rather than reflecting formal definitions of assessment, students' views are grounded in their ongoing participation in classroom activities and in the meanings they attribute to teachers' actions.

Research shows that students interpret assessment differently depending on how it is enacted. When assessment is experienced as supportive, relevant, and oriented toward improvement, students tend to report higher engagement and more positive emotional responses (Birenbaum et al., 2006). Conversely, when it is experienced as primarily grade-focused or disconnected from learning, students tend to report reduced motivation, lower ownership of learning, and a sense of irrelevance (Brown, 2011; Dann, 2014). These perceptions represent not only reactions to specific tasks but interpretations of the broader learning environment.

The classroom motivational climate provides the context in which such interpretations occur. Climate reflects the interpersonal tone, expectations, and participation structures established by the teacher and interpreted by students (Ames, 1992; Patrick et al., 2007). Students' sense of support, fairness, and autonomy depends on how teachers organize activities, respond to performance, and communicate expectations (Alonso-Tapia & Fernández-Heredia, 2008; Reeve, 2006). Importantly, climate is not a background variable, but a product of teacher behaviour as perceived by students, linking pedagogical actions to motivational responses (Alonso-Tapia & Fernández-Heredia, 2008).

Assessment practices are a central component of this climate because evaluation routines repeatedly signal how competence is defined and how students are positioned within the classroom (Brookhart, 2011; Cizek, 2010). Practices emphasizing feedback, progress, and clear expectations tend to be interpreted as supportive, whereas heavy reliance on grading, ranking, or opaque criteria may be experienced as controlling or discouraging (Black & Wiliam, 2009; McMillan, 2013; Wiliam, 2011). In this sense, classroom climate and assessment are closely intertwined: students often infer the quality of their relationship with the teacher from how they are evaluated.

This interpretive process is particularly salient in physical education. Because performance is visible and often socially comparative, students may experience assessment as motivating and collaborative when effort and improvement are emphasized, or as evaluative and externally imposed when attention is focused primarily on performance outcomes (Hay & Penney, 2013; López-Pastor et al., 2013). The same assessment activity may therefore convey different meanings depending on how it is structured and communicated. These concerns are also reflected in international discussions within the field, such as the AIESEP (2020) position statement on assessment in physical education, which emphasises the importance of aligning assessment practices with learning and student development. Beyond the visibility of performance, assessment in physical education has been theorised as an inherently relational practice, in which the meaning of evaluation is co-constructed between teachers and students rather than residing in the assessment task itself (Hay & Penney, 2013). From this perspective, students' capacity to recognize and act on the purposes of assessment, often discussed as students' assessment literacy, is not a fixed individual attribute but develops through the evaluative interactions and shared expectations that

characterize each class (Hay & Penney, 2013; López-Pastor et al., 2013). This relational and literacy-oriented view reinforces our central premise: if the meaning of assessment is built within classroom interaction, students taught by different teachers may interpret the same evaluative practices in systematically different ways.

Taken together, these findings suggest that assessment operates within classrooms as an interpreted experience. Students construct meanings about fairness, usefulness, and personal value through repeated exposure to assessment routines and interactions with their teachers. If teachers differ in how they approach and communicate assessment, classrooms are likely to differ systematically in how students perceive both assessment and the motivational climate. This expectation provides the basis for examining variation between teachers and linking teacher-level characteristics to student-reported perceptions.

If assessment is experienced by students through ongoing classroom interactions, its meaning is likely not identical across classrooms. Even when teachers operate within the same curriculum and institutional context, they differ in how they communicate expectations, provide feedback, and structure participation (Black & Wiliam, 2009; Brookhart, 2011). Because students repeatedly encounter these routines, those being taught by the same teacher within the same classroom are likely to develop common perceptions, while differing from students taught by other teachers (Brookhart & Durkin, 2003; Brown & Hirschfeld, 2008; Struyven et al., 2005).

From this perspective, the classroom becomes a meaningful unit of analysis. Students in a given class are exposed to similar interactional patterns and evaluative practices, and therefore their perceptions of assessment and classroom climate should show a degree of similarity. Research on classroom motivational climate indicates that students interpret teaching practices collectively and construct shared views of support, expectations, and fairness within the same classroom (Alonso-Tapia & Fernández-Heredia, 2008; Ames, 1992; Patrick et al., 2007; Reeve, 2006). In turn, differences in teachers' approaches to assessment should be reflected in systematic differences between classes.

This reasoning implies that student perceptions are not solely individual attitudes, but somewhat collective experiences shaped by classroom processes. A multilevel approach therefore allows identification of variation attributable to the teacher-classroom context and facilitates the examination of whether teacher characteristics are associated with classroom-level differences (Brookhart, 2011; Cizek, 2010).

3. Teacher characteristics and their role in classroom dynamics

Teachers' demographic characteristics, such as age and gender, are increasingly recognized as relevant factors in shaping classroom dynamics, pedagogical practices, and student outcomes (Klassen & Tze, 2014; Spilt et al., 2011). Because students experience assessment through ongoing classroom interactions, such characteristics may also be associated with how evaluative practices are interpreted by students.

Teacher age, which often correlates with professional experience and socialization into institutional norms, has been linked to variation in instructional and assessment approaches. Some studies suggest that more experienced or older teachers are more likely to rely on traditional or summative-oriented practices, whereas younger teachers may be more open to formative or student-centered approaches (Koh, 2011; McMillan, 2013). At the same time, students may perceive teachers differently based on age-related interactional styles, such as approachability, flexibility, or perceived authority, thus shaping their experience of the motivational climate (Nurmi, 2012).

Gender has likewise been examined in relation to classroom processes, although findings are context-dependent. Some evidence indicates that female teachers more frequently adopt student-centered and mastery-oriented instructional strategies, which may affect how students perceive fairness, support, and encouragement (Retelsdorf et al.,

2010; Rubie-Davies et al., 2012; Wolters & Daugherty, 2007). However, associations between teacher gender and assessment practices remain inconsistent (Harris & Brown, 2009).

In the present study, teacher age and gender are included as level-2 predictors in the multilevel models to examine whether they explain differences in students' perceptions of assessment and classroom motivational climate.

4. Aim and research questions

The aim of this study is to investigate, using a multilevel approach, whether students' perceptions of assessment and classroom motivational climate vary across classes and whether teacher-level characteristics, including assessment conceptions, age, and gender, are associated with this variation. Two research questions (RQ) guide this investigation:

RQ1. Do students' perceptions of assessment and classroom motivational climate vary across classes (i.e., between teachers) in secondary physical education?

RQ2. Do teachers' assessment conceptions, age, and gender account for differences in students' perceptions of assessment and classroom motivational climate in secondary physical education?

5. Method

5.1. Participants

The study involved a total of 989 secondary school students from Spain and their physical education teachers ($n = 17$), selected using convenience sampling. Participants came from 10 schools (6 public and 4 private) located in five different districts of the Madrid metropolitan area. This distribution implies high variability in the socio-economic background of our participants, both because of the type of school (i.e., public and private) and their geographical distribution. Among the students, 454 (46.0%) identified as male, 448 (45.3%) as female, 20 (2.0%) selected "other", 66 (6.7%) preferred not to disclose their gender, and one student did not respond to the gender question. The average age of the student participants was 14.6 years, with ages ranging from 13 to 17.

The teacher sample consisted of 17 participants, including 12 (70.6%) males and 5 (29.4%) females. The average age of the teachers was 36.6 years, with a range between 23 and 57 years. Their overall teaching experience averaged 9.4 years (ranging from 3 months to 30 years), while their experience at their current school averaged 6.6 years (ranging from 3 months to 23 years).

5.2. Instruments

Four questionnaires were used in this study: two completed by teachers and two by students. As part of their questionnaires, teachers indicated their gender (*female* and *male*) and their age, as well as their perceived formative orientation (on a scale from 0 to 10), and these variables were used in the analysis in this study.

Before describing each instrument in detail, it is important to note that we conducted exploratory factor analyses (EFA) only for the student questionnaires. This decision was based on sample size considerations: the student sample ($n = 989$) was sufficiently large to support reliable factor analytic results, whereas the teacher sample ($n = 17$) was too small to yield meaningful outcomes. Consequently, we only report EFA results for the student scales in the sections below. Similarly, internal consistency (Cronbach's alpha) was estimated for the student scales, while for the teacher scales, we report reliability estimates from their original validation studies.

5.2.1. Battery of Teachers' Motivational Variables Related to Students (B-TMV-S)

This questionnaire was initially developed by Garazo and Alonso-Tapia (2016) and then validated by Alonso-Tapia & Ruiz-Díaz, 2022, to measure teachers' motivational orientation. In total, it consists of 27 Likert-type items, with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), and it includes three subscales: teacher focus on student grades (9 items; $\alpha = .76$); teacher focus on student emotional needs (10 items; $\alpha = .84$), and teacher expectancies on managing students (8 items; $\alpha = .66$), which is split into two subscales: i) teacher self-efficacy expectancies and ii) teacher control expectancies (Alonso-Tapia & Ruiz-Díaz, 2022).

5.2.2. Qualitative Model of Conceptions of Assessment (QMCoA)

This questionnaire, developed by Remesal and Brown (2015), consists of 34 items and measures two main teachers' conceptions regarding the *purpose* of assessment (formative regulation and societal control) and four conceptions regarding the *focus* of assessment (teaching, learning, certification, and accountability). The questionnaire uses a six-point Likert-type scale, including four positive and two negative response options. Cronbach's alpha coefficients reported by Brown & Remesal, 2017 were .83 for formative regulation, .74 for societal control, .57 for teaching, .51 for learning, .57 for certification, and .66 for accountability.

In this study, only the two purpose-related subscales, formative regulation (example item: "Each assessment occasion is also a new opportunity for learning") and societal control (example item: "If there was not learning assessment, it would be impossible to motivate students to learn"), were used, as together they encompass all items included in the four focus-related subscales (teaching, learning, certification, and accountability). This decision aimed to keep the models as parsimonious as possible, while still capturing the full conceptual range of teachers' assessment conceptions represented in the original instrument.

5.2.3. Students Conceptions of Assessment (SCoA-VI)

This questionnaire (Brown, 2011) consists of 33 items and measures students' conceptions of assessment in four inter-related constructs: (1) assessment improves learning and teaching (improvement), (2) assessment relates to external factors (external attributions), (3) assessment has affective and social benefits (affect/social benefits), and (4) assessment is irrelevant (irrelevance). The original questionnaire presented Cronbach's alphas ranging from .82 to .94 (Brown et al., 2009). For the purposes of this study, this questionnaire was translated to Spanish.

To examine the suitability of the instrument for our sample, we conducted an EFA. The results confirmed the original four-factor structure, as all items loaded reasonably well onto their intended factors. While a few items showed secondary loadings on other factors, these were lower than the primary loadings, thus not warranting any changes. Therefore, the original four-factor structure was retained. Reliability analyses also yielded acceptable results, with Cronbach's alpha coefficients for the four subscales ranging from .75 to .85.

5.2.4. Classroom Motivational Climate Questionnaire (CMC-Q)

This questionnaire (Alonso-Tapia & Fernández-Heredia, 2008) was developed to assess how students perceive the motivational climate of their classroom. It measures 16 specific teaching strategies or patterns that could affect student motivation to learn. Each strategy/pattern is represented by two items, one positively and one negatively worded, resulting in a total of 32 items. Reliability analysis performed in the original study showed very high Cronbach's alpha coefficients among secondary and high school students ($\alpha = .93$, respectively).

To examine the suitability of the instrument for our sample, we conducted an EFA. The results showed that all items had their highest loading on the first factor, except for item 26 that fell slightly below the conventional loading threshold of .30 (loading = .27) and exhibited moderate cross-loadings on two additional factors. Given that this was

the only item that did not meet the threshold, we conducted a reliability analysis to inform the decision regarding its retention. The full scale demonstrated a very high Cronbach's alpha coefficient ($\alpha = .94$). Removing item 26 led to a negligible increase by one unit at the third decimal place, suggesting that its exclusion would not meaningfully improve the psychometric properties of the scale. Therefore, the original 32-item structure was retained.

5.3. Procedure

The study was approved by the Ethical Committees of Universidad de Deusto and Universidad Politécnica de Madrid (reference: ETK-22/21-22). After this approval, the last author contacted physical education teachers who served as practicum tutors for university students. These teachers were invited to participate in the research and to facilitate the involvement of their students as well. Before the questionnaires were distributed, informed consent was obtained from all participants. Teachers signed consent forms directly, while parents and guardians provided written authorization for students' participation. Participation was voluntary and no incentives were offered.

Data collection was organized to respect the availability of each teacher. A member of the research team visited each school to conduct in-person data collection during regularly scheduled physical education classes. These sessions ensured that students could complete the questionnaires in a familiar setting and that any doubts about the items could be clarified by the researcher. Student questionnaires were primarily administered using a paper-and-pencil format. However, in one school, students completed the questionnaires online. All teacher responses were collected online. The data collection took place between November and December 2022.

To ensure anonymity and to link student and teacher data for analytical purposes, a coding system was established. Each school, teacher, and student were assigned a unique code, allowing for matched data while preserving confidentiality. No personally identifiable information was collected.

5.4. Data analyses

We first conducted descriptive analyses using SPSS 28 (IBM Corp., 2021). To examine how teacher-level characteristics may be related to students' perceptions of assessment and classroom motivational climate, we then employed a two-level hierarchical modeling strategy, treating students (Level 1) as nested within teachers (Level 2). This multilevel framework was selected for both theoretical and statistical reasons. Theoretically, the research aims to capture how teacher assessment-related conceptions and demographic characteristics related to the broader assessment culture experienced by students in their classrooms. Statistically, the use of multilevel modeling accounts for the non-independence of observations within teacher groups and enables the partitioning of variance across levels (Snijders & Bosker, 2012).

A series of two-level linear regression models with random intercepts were estimated using maximum likelihood with robust standard errors, with listwise deletion applied to missing data. Given the low proportion of missing cases, this approach for handling missing data did not introduce significant bias or affect the results. Separate models were constructed for each of the five student-reported outcomes: perceived classroom motivational climate (CMC total) and the four dimensions of assessment perceptions: (i) SCoA improvement, (ii) SCoA irrelevance, (iii) SCoA affect/social benefit, and (iv) SCoA external attributions. Statistical significance was set at $p < .05$ and the modelling was conducted in Mplus 8 (Muthén & Muthén, 2017).

The multilevel modelling was conducted in two steps. First, null models (i.e., models with no predictors) were estimated to determine the proportion of variance in each outcome attributable to between-teacher differences, as indicated by the intraclass correlation coefficient (ICC). These null models also served as a baseline for comparing the full

models, allowing the evaluation of the extent to which the inclusion of predictor variables contributed to the prediction of each outcome. Second, teacher-level predictors were added to the models to determine whether they predicted each of the student-reported outcomes.

6. Results

Table 1 presents the means and standard deviations (SD) for each of the five outcomes: perceived classroom motivational climate (CMC total), the four dimensions of assessment perceptions: (i) SCoA improvement, (ii) SCoA irrelevance, (iii) SCoA affect/social benefit, and (iv) SCoA external attributions, and the continuous predictors: teacher motivational variables (teacher focus on student grades, teacher focus on student emotional needs, teacher self-efficacy expectancies, and teacher control expectancies), conceptions of assessment as societal control, conceptions of assessment as formative, and a global 0–10 rating of how formative the teacher perceives their own practice.

6.1. RQ1. Do students' perceptions of assessment and classroom motivational climate vary across classes (i.e., between teachers) in secondary physical education?

To examine the degree to which students' perceptions were shaped by students' nesting within teachers, null models were estimated for each outcome. These models included no predictors and allowed for the estimation of the ICCs, quantifying the proportion of total variance in each of the student-reported outcomes attributable to differences between teachers.

Table 2 presents the ICCs from the null models. ICCs ranged from .071 in model 5 to .189 in model 1, indicating that between 7.1% and 18.9% of the variance in students' responses was attributable to their nesting within teachers. These values provided justification for the use of multilevel models in the analyses, confirming that students' perceptions of both assessment and classroom motivational climate are likely shaped by the classroom context formed by their teachers.

6.2. RQ2. Do teachers' assessment conceptions, age, and gender explain differences in students' perceptions of assessment and classroom motivational climate in secondary physical education?

To address this question, five two-level linear regression models were estimated, each predicting a different student-reported outcome. Teacher-level predictors included age, gender, and seven composite measures from three sources: four motivational orientation scales (teacher focus on student grades, teacher focus on student emotional needs, teacher self-efficacy expectancies, and teacher control expectancies), two conceptions of assessment scales (as societal control and as formative), and a global 0–10 rating of how formative the teacher perceives their own practice. Table 3 summarizes the model results providing the conditional ICC and the teacher-level variance explained for each model as well as the unstandardized coefficients (B) and

Table 2

Intraclass correlation coefficients (ICCs) from null models.

Variable	ICC
CMC total	0.189
SCoA improvement	0.122
SCoA irrelevance	0.187
SCoA affect/social benefit	0.166
SCoA external attributions	0.071

Note. Model 1: n (students) = 968; Model 2: n (students) = 963; Model 3: n (students) = 967; Model 4: n (students) = 981; Model 5: n (students) = 971.

standardized coefficients (β), along with their standard errors (SE), and the statistical significance (p -value) for each predictor variable in each model.

In the first model focusing on students' perceptions of classroom motivational climate (*CMC total*), two teacher-level predictors were statistically significant. Teacher focus on student grades was negatively associated with students' perceptions of classroom motivational climate ($\beta = -0.358$, $SE = 0.137$), indicating that students perceived a less positive classroom climate when teachers reported focusing more on student grades. Teacher age was also negatively related to the outcome ($\beta = -0.695$, $SE = 0.266$); the older the teacher, the lower the students' perceptions of classroom motivational climate. The model explained 70.8% of the between-teacher variance.

In the second model focusing on students' perceptions of assessment as improvement-oriented (*SCoA improvement*), teacher focus on student grades ($\beta = -0.446$, $SE = 0.151$) and their age ($\beta = -0.881$, $SE = 0.227$) were statistically significant and negative predictors of the outcome. Students perceived assessment as more improvement-oriented when teachers reported focusing less on student grades or were younger. The model explained 79.0% of the between-teacher variance.

In the third model focusing on students' perceptions of assessment as irrelevant (*SCoA irrelevance*), teachers' global self-rating of formative orientation ($\beta = 1.007$, $SE = 0.352$) and their age ($\beta = 0.938$, $SE = 0.236$) were statistically significant and positively associated with students' perceiving assessment as more irrelevant. This suggests that students viewed assessment as more irrelevant to their learning when teachers rated their practice as more formative or were older. The model explained 78.3% of the between-teacher variance.

In the fourth model focusing on students' perceptions of affective or social benefits of assessment (*SCoA affect/social benefit*), teacher focus on student grades ($\beta = -0.426$, $SE = 0.132$), teacher self-efficacy expectancies ($\beta = -0.683$, $SE = 0.190$), teacher global self-rating of formative orientation ($\beta = -1.240$, $SE = 0.385$), and their age ($\beta = -1.008$, $SE = 0.284$) were statistically significant and negative predictors of the outcome; students taught by teachers who reported a greater focus on student grades, higher self-efficacy expectancies, a stronger formative assessment orientation, and who were older tended to perceive fewer affective or social benefits of assessment. The model explained 81.2% of the between-teacher variance.

In the fifth model focusing on students' perceptions of assessment outcomes being determined by factors outside their control (*SCoA external attributions*), teacher conceptions of assessment as societal control were statistically significantly and positively related to the outcome ($\beta = 0.620$, $SE = 0.182$), suggesting that students were more likely to attribute assessment outcomes to external or uncontrollable factors when their teacher endorsed societal control views. Teacher global self-rating of formative orientation ($\beta = -0.748$, $SE = 0.323$) and age ($\beta = -1.028$, $SE = 0.247$), again, showed statistically significant and negative associations with the outcome, with students taught by teachers who identified more strongly as formative assessors and older teachers tending to report lower levels of perceived external attributions. The model explained 99.4% of the between-teacher variance.

Taken together, these results highlight teacher age and perceived formative orientation as the most consistent predictors across models,

Table 1

Means and standard deviations for outcomes and continuous predictors.

Variable	M	SD
CMC total	115.45	21.75
SCoA improvement	3.46	0.70
SCoA irrelevance	2.57	0.78
SCoA affect/social benefit	3.41	0.67
SCoA external attributions	2.71	0.78
Teacher focus on student grades	1.83	0.42
Teacher focus on student emotional needs	4.39	0.40
Teacher self-efficacy expectancies	3.80	0.36
Teacher control expectancies	2.40	0.58
Teacher conceptions of assessment as societal control	2.49	0.34
Teacher conceptions of assessment as formative	4.04	0.29
Teacher global self-rating of formative orientation	7.14	1.85

Table 3

Two-level linear regression models of perceived classroom motivational climate and the four dimensions of assessment perceptions.

CMC total					
ICC	0.191				
R ² – teacher-level (%)	70.8				
	B	SE	β	SE	p-value
Teacher focus on student grades	−7.952	2.978	−0.358	0.137	0.008
Teacher focus on student emotional needs	1.609	6.548	0.071	0.288	0.806
Teacher self-efficacy expectancies	2.889	5.523	0.110	0.216	0.601
Teacher control expectancies	−3.143	6.637	−0.202	0.419	0.636
Teacher conceptions of assessment as societal control	5.844	5.302	0.253	0.226	0.270
Teacher conceptions of assessment as formative	0.758	13.183	0.021	0.361	0.954
Teacher global self-rating of formative orientation	−2.970	1.897	−0.624	0.387	0.117
Teacher age	−0.724	0.306	−0.695	0.266	0.018
Teacher gender (female)	0.674	3.210	0.069	0.329	0.834
SCoA improvement					
ICC	0.129				
R ² – teacher-level (%)	79.0				
	B	SE	β	SE	p-value
Teacher focus on student grades	−0.261	0.079	−0.446	0.151	0.001
Teacher focus on student emotional needs	0.078	0.163	0.131	0.272	0.633
Teacher self-efficacy expectancies	−0.130	0.129	−0.187	0.187	0.316
Teacher control expectancies	−0.180	0.124	−0.440	0.312	0.148
Teacher conceptions of assessment as societal control	0.182	0.128	0.301	0.218	0.155
Teacher conceptions of assessment as formative	0.197	0.346	0.205	0.369	0.569
Teacher global self-rating of formative orientation	−0.093	0.048	−0.742	0.425	0.054
Teacher age	−0.024	0.007	−0.881	0.227	0.000
Teacher gender (female)	0.003	0.072	0.014	0.281	0.961
SCoA irrelevance					
ICC	0.193				
R ² – teacher-level (%)	78.3				
	B	SE	β	SE	p-value
Teacher focus on student grades	0.158	0.093	0.196	0.115	0.088
Teacher focus on student emotional needs	−0.023	0.214	−0.029	0.262	0.913
Teacher self-efficacy expectancies	0.071	0.153	0.075	0.161	0.641
Teacher control expectancies	0.275	0.163	0.489	0.298	0.091
Teacher conceptions of assessment as societal control	−0.298	0.157	−0.358	0.192	0.058
Teacher conceptions of assessment as formative	−0.534	0.409	−0.405	0.320	0.192
Teacher global self-rating of formative orientation	0.173	0.053	1.007	0.352	0.001
Teacher age	0.035	0.008	0.938	0.236	0.000
Teacher gender (female)	−0.035	0.094	−0.099	0.269	0.710
SCoA affect/social benefit					

Table 3 (continued)

ICC	0.164				
R ² – teacher-level (%)	81.2				
	B	SE	β	SE	p-value
Teacher focus on student grades	−0.273	0.075	−0.426	0.132	0.000
Teacher focus on student emotional needs	0.130	0.200	0.200	0.298	0.516
Teacher self-efficacy expectancies	−0.518	0.126	−0.683	0.190	0.000
Teacher control expectancies	−0.245	0.179	−0.546	0.398	0.172
Teacher conceptions of assessment as societal control	0.191	0.149	0.288	0.223	0.201
Teacher conceptions of assessment as formative	0.555	0.347	0.527	0.351	0.110
Teacher global self-rating of formative orientation	−0.170	0.049	−1.240	0.385	0.001
Teacher age	−0.030	0.008	−1.008	0.284	0.000
Teacher gender (female)	−0.135	0.094	−0.483	0.356	0.148
SCoA external attributions					
ICC	0.104				
R ² – teacher-level (%)	99.4				
	B	SE	β	SE	p-value
Teacher focus on student grades	0.024	0.068	0.040	0.117	0.728
Teacher focus on student emotional needs	0.199	0.143	0.334	0.238	0.165
Teacher self-efficacy expectancies	−0.078	0.081	−0.112	0.118	0.335
Teacher control expectancies	−0.207	0.122	−0.506	0.324	0.089
Teacher conceptions of assessment as societal control	0.377	0.077	0.620	0.182	0.000
Teacher conceptions of assessment as formative	0.180	0.289	0.187	0.308	0.534
Teacher global self-rating of formative orientation	−0.094	0.034	−0.748	0.323	0.006
Teacher age	−0.028	0.005	−1.028	0.247	0.000
Teacher gender (female)	0.084	0.056	0.329	0.230	0.133

Notes. Reference category of categorical variables in parentheses. Some standardized coefficients exceeded an absolute value of 1. This can occur in models with correlated predictors and relatively small Level-2 sample sizes and may reflect multicollinearity or suppression effects. Therefore, individual standardized coefficients should be interpreted with caution.

although in some cases the associations were unexpected or counterintuitive, followed by teacher focus on grades. Notably, teacher gender and the remaining teacher-level measures did not emerge as consistent predictors across the models.

In interpreting the model results, it is important to consider the limitations associated with the employed scales, as described in the *Method* section. In addition, the model fit indices indicated that, in all cases, the full models did not demonstrate a better fit than their corresponding null models. The combination of high proportions of explained teacher-level variance and no improvement in fit over the corresponding null models likely reflects the small number of teacher-level units ($n = 17$) relative to the number of teacher-level predictors included in the models. Consequently, the R^2 estimates should be interpreted with caution, as they may overestimate the explanatory power of the predictors. This issue is particularly evident in the fifth model (SCoA external attributions), where the highest R^2 (99.4%) was observed alongside the lowest ICC (7.1%). These considerations suggest that alternative model specifications may offer a more informative understanding of the data, particularly in light of the constraints imposed by the scales.

7. Discussion

Our findings suggest that classroom assessment is experienced by students not merely as a set of instructional procedures but as an interpreted classroom experience. Students taught by different teachers reported systematically different perceptions of assessment and classroom motivational climate, indicating that the meaning of assessment is at least partly constructed within each classroom context. At the same time, teachers' own conceptions of assessment were not necessarily aligned with their students' related perceptions. In particular, stronger self-reported formative orientation among teachers did not correspond to more positive student perceptions and, in some dimensions, was associated with less favourable ones. Taken together, these results indicate that assessment does not operate as a direct transmission of teachers' intentions. Instead, its effects depend on how evaluative practices are enacted and interpreted within ongoing teacher-student interactions. The study therefore highlights assessment as a relational and interpretive process in which classroom dynamics shape how evaluation is understood and experienced by students.

7.1. Assessment as an interpreted classroom experience

The multilevel analyses indicated consistent between-teacher variation in all student-reported outcomes. The ICCs showed that a substantial proportion of variance in students' perceptions was attributable to the class level, suggesting that students within the same class tended to share similar perceptions of assessment and motivational climate. Rather than reflecting individual attitudes alone, students' responses appear to represent shared interpretations of evaluative practices within the classroom. This finding supports the view that assessment operates as a social and relational process embedded in classroom interactions. Students exposed to the same teacher repeatedly encounter similar routines of feedback, expectations, and evaluative communication, and therefore develop convergent understandings of what assessment means in that context. Previous research on classroom climate has similarly shown that students form shared perceptions of support, structure, and fairness within classrooms (Alonso-Tapia & Fernández-Heredia, 2008; Patrick et al., 2007). Our results extend this idea to assessment, indicating that evaluative experiences are also collectively constructed.

The presence of class-level variation has theoretical implications that extend beyond the statistical justification for using multilevel modelling. It indicates that assessment is not experienced uniformly across instructional settings but acquires distinct meanings in different classrooms. Even when curricular objectives are comparable, teachers' ways of structuring evaluation appear to create classroom-specific assessment environments (Fraser, 2012). Consequently, understanding students' perceptions of assessment requires considering the teacher-classroom context in which those perceptions are formed.

7.2. Teacher intentions and students' experiences of assessment

The analyses showed that teachers' formative orientation was not consistently associated with more positive student perceptions of assessment. For some outcomes, a stronger self-reported formative orientation was associated with less favourable student perceptions. Rather than contradicting formative assessment theory, this pattern likely suggests a misalignment between teachers' intentions and how assessment is experienced by students. Previous research has shown that classroom assessment functions not only as an instructional technique but also as a relational and motivational signal conveying expectations about competence and participation (Black & Wiliam, 2009; Brookhart, 2011). Our findings indicate that these signals may be interpreted by students in ways that differ from teachers' intended meanings. An exception to this pattern was found for external attributions, where a stronger self-reported formative orientation predicted lower levels of attributing assessment outcomes to factors beyond students' control.

This pattern is consistent with the expectation that formative practice help students view assessment as relevant to their own learning. It should also be noted that teachers' conceptions of assessment as formative, a construct conceptually similar to their global self-rating of formative orientation, did not statistically significantly predict any of the five student-reported outcomes. This discrepancy suggests that the associations observed for the global self-rating should be interpreted cautiously. As a single-item self-evaluation, the measure may be more susceptible to self-enhancement biases than a more comprehensive measurement of formative conceptions.

Research on teacher beliefs has consistently documented a gap between espoused conceptions and enacted practices (Bennett, 2011; Harris & Brown, 2009). Teachers may endorse formative purposes while still working within routines that emphasise grading, performance comparison, or limited opportunities to use feedback. Students, however, encounter assessment through concrete classroom events (e.g., how criteria are explained, how feedback is provided, how evaluation affects participation). Consequently, students' perceptions are shaped less by teachers' reported conceptions and more by the assessment practices they repeatedly experience.

This interpretation aligns with work on conceptions of assessment suggesting that the meaning of assessment depends on its communication and enactment (Brown & Hirschfeld, 2008; Brown, 2008). What teachers consider formative may not be recognised as such by students if expectations are unclear, feedback is difficult to use, or evaluative consequences remain salient. In this sense, assessment operates as a communicative and interpretive process (Brookhart, 2011; Cizek, 2010), and its motivational consequences depend on how students perceive its fairness, usefulness, and relevance to learning (Birenbaum et al., 2006; Struyven et al., 2005).

The pattern observed for teachers' focus on grades supports this interpretation. A stronger teacher focus on grading was associated with lower student-reported perceptions of classroom motivational climate, assessment as improvement-oriented, and assessment as affectively/socially beneficial. Prior research has shown that when evaluation is experienced as judgement rather than guidance, students report lower engagement and more performance-focused responses (Alonso-Tapia & Fernández-Heredia, 2008; Patrick et al., 2007). Thus, students appear sensitive to the evaluative messages conveyed by assessment practices, particularly the extent to which they emphasise learning progress versus comparison.

Taken together, the findings suggest that assessment meaning is constructed within classroom interaction rather than transmitted directly from teachers' beliefs to students' perceptions. Teachers' conceptions may guide their intentions, but students respond to the assessment environment they experience. The results therefore point to a translation process between pedagogical intention and student interpretation, in which classroom communication mediates the motivational and relational consequences of assessment (Brookhart, 2011; Fraser, 2012).

7.3. Teacher demographic characteristics and student perceptions

Teacher age was negatively associated with student perceptions of classroom motivational climate, improvement-oriented assessment, the affective and social benefits of assessment, and external attributions of assessment outcomes, but positively associated with perceptions of assessment irrelevance. These results should not be interpreted as effects of age itself but as indicators of differences in interactional and instructional patterns that may accompany professional experience. Prior research has suggested that teaching experience and professional socialisation can influence pedagogical routines and assessment practices (Koh, 2011; McMillan, 2013). More experienced teachers may rely on established procedures that emphasise evaluation and classroom control, which students may interpret as less supportive or less responsive to their learning needs.

From a classroom interaction perspective, students respond to how teachers communicate expectations, feedback, and evaluation. Studies of classroom climate show that perceptions of support, fairness, and clarity depend strongly on teacher-student interactions rather than on formal instructional intentions (Alonso-Tapia & Fernández-Heredia, 2008; Patrick et al., 2007). Age may therefore function as a contextual factor influencing communication style, authority relationships, or responsiveness, which, in turn, shape how assessment is interpreted by students.

Teacher gender did not predict any student-reported outcomes. This is consistent with previous work showing limited and context-dependent associations between teacher gender and student perceptions (Harris and Brown, 2009; Rubie-Davies et al., 2012; Spilt et al., 2011).

7.4. Implications for assessment practice and teacher education

Our findings have implications for how assessment is understood in educational practice. Teacher assessment literacy is often framed in terms of designing appropriate tasks, using formative strategies, and applying fair grading procedures. However, the present results suggest that assessment effectiveness also depends on how students interpret evaluative practices. Even when teachers report formative intentions, students may not experience assessment as supportive if expectations, feedback, or purposes are not clearly communicated. Assessment literacy, therefore, involves not only selecting suitable assessment methods but also ensuring that the meaning of assessment is transparent and recognizable to learners.

This has consequences for teacher education and professional development. Programs that promote formative assessment frequently emphasise principles and techniques, yet teachers may need support in understanding how their practices are perceived by students. Encouraging teachers to gather and reflect on student perception data may help identify discrepancies between pedagogical intentions and classroom experience. Such reflection could strengthen alignment between feedback, grading practices, and students' understanding of learning goals.

The results also suggest that attention should be given to assessment communication. How criteria are explained, how feedback is framed, and how grades are integrated into instruction appear central to students' motivational responses (Brookhart, 2011; Cizek, 2010). In subjects such as physical education, where performance is visible and socially contextualised, clarity and transparency may be especially important for maintaining a positive classroom climate and positive student perceptions. This is consistent with sociocultural accounts of assessment in physical education, which frame evaluation as a relational practice whose meaning is negotiated between teachers and students rather than transmitted through the assessment task alone (Hay & Penney, 2013). Read through this lens, our finding that stronger self-reported formative orientation did not translate into more positive student perceptions points to the value of developing students' assessment literacy, that is, their capacity to recognise the purposes and worth of the assessment they experience, alongside teachers' own assessment practices.

More broadly, the study indicates that evaluation should be considered a relational process rather than solely a measurement procedure. Assessment practices contribute to how students interpret competence, effort, and fairness within the classroom environment. For this reason, improving assessment practice may require not only better tools or procedures but also greater attention to how assessment experiences are constructed through everyday teacher-student interactions.

8. Limitations

Several limitations should be considered when interpreting the findings. First, the cross-sectional design does not allow causal inferences. Although associations were found between teacher characteristics and student perceptions, the direction and mechanisms of these

relationships cannot be determined. Second, both teacher and student variables were measured through self-report instruments. While the measures were validated, responses may be influenced by perception biases or incomplete awareness of classroom practices. Third, the number of teacher participants was limited ($n = 17$), restricting the statistical power and the complexity of models that could be estimated (e.g., interactions, mediation). Finally, the study was conducted within a specific context (i.e., Spanish secondary physical education), which may limit the generalizability of results to other subjects or education systems. Collectively, these limitations underscore the exploratory nature of the study and highlight the need for replication with larger (particularly at the teacher level) and more diverse samples. Future research of this kind would enable more rigorous measurement of the constructs under investigation and provide stronger evidence before firmer conclusions can be established.

9. Conclusion

This study examined how teacher-level characteristics relate to students' perceptions of assessment and classroom motivational climate in secondary physical education. Our findings indicate that assessment is not experienced solely as a technical or instructional practice but as a classroom process interpreted by students within ongoing teacher-student interactions. Students taught by different teachers developed systematically different understandings of assessment, and these understandings were not directly aligned with teachers' reported intentions.

The results therefore suggest that what assessment means to students depends not only on what teachers intend assessment to accomplish but on how its purposes and consequences are communicated and recognised by learners. Assessment practices shape classroom relationships, expectations, and motivation through the meanings students attribute to them. Consistent with sociocultural views of assessment in physical education, evaluation operated here as a relational process in which meaning was negotiated through everyday interactions within the classroom rather than transmitted directly from teachers' conceptions to students' perceptions (Hay & Penney, 2013). Understanding these dynamics therefore requires attention not only to what teachers intend but to the assessment literacy that students develop as they make sense of repeated evaluative experiences.

By showing that students' perceptions do not necessarily mirror teachers' conceptions, the study shifts attention from assessment design alone to assessment interpretation. Improving assessment may thus require not only better methods, but greater attention to how evaluative practices are enacted and understood in classrooms.

For initial teacher education and professional learning, these findings indicate that preparing teachers to assess well involves more than mastering formative techniques. It also requires helping teachers anticipate how their assessment practices are likely to be interpreted by students, and supporting them in gathering and reflecting on student perception data as a routine part of their practice. Such reflection may be especially valuable in physical education, where the visibility and social comparability of performance heighten the relational stakes of evaluation.

Finally, the limitations of this study point to a clear agenda for future research. Replication with larger teacher samples would allow more complex modelling, including the examination of interaction and mediation effects. Longitudinal and observational designs that combine student perception data with direct evidence of enacted assessment practices would help clarify how the misalignment we observed develops and whether it can be reduced. Extending this work to other subjects and education systems would further establish how far the relational and interpretive dynamics identified here are specific to physical education or common across different assessment contexts.

CRediT authorship contribution statement

Ernesto Panadero: Writing – review & editing, Writing – original draft, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Vasilliki Pitsia:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Javier Fernández Ruiz:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Daniel García-Pérez:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Rodrigo Pardo:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization.

Funding

(1) Spanish National R + D call from the Ministerio de Ciencia, Innovación y Universidades (Generación del conocimiento 2020), Reference number: PID2019–108982GB-I00. (2) Basque Government Call for Grants to support the activities of research groups of the Basque University System (2022–2025) project reference IT1624–22.

Declaration of Competing Interest

The authors declare to not having any conflict of interest regarding this manuscript. This research has been approved by the ethics committees from Comité de ética de la investigación, Universidad de Deusto and Universidad Politécnica de Madrid. Reference: ETK-22/21-22. PI: Ernesto Panadero.

References

- AIESEP. (2020). *AIESEP position statement on physical education assessment*. <https://aiesep.org/scientific-meetings/position-statements/>.
- Alonso-Tapia, J., & Fernández-Heredia, B. (2008). Development and initial validation of the Classroom Motivational Climate Questionnaire (CMCQ). *Psicothema*, 20(4), 883–889.
- Alonso-Tapia, J., & Ruiz-Díaz, M. (2022). School climate and teachers motivational variables: Effects on teacher satisfaction and classroom motivational climate perceived by middle school students: A cross-cultural study. *Psicología Educativa*, 28(2), 151–163. <https://doi.org/10.5093/psed2022a4>
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271. <https://doi.org/10.1037/0022-0663.84.3.261>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R., & Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, 1(1), 61–67. <https://doi.org/10.1016/j.edurev.2006.01.001>
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27–54. https://doi.org/10.1207/S15324818AME1601_2
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. Nova Science Publishers.
- Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: A review of the Students' Conceptions of Assessment inventory. *Educational Psychology*, 31(6), 731–748. <https://doi.org/10.1080/01443410.2011.599836>
- Brown, G. T. L. (2017). *Assessment of student achievement* (1st ed.). Routledge. <https://doi.org/10.4324/9781315162058>
- Brown, T. D., Barnes, M., & Finefer-Rosenbluh, I. (2024). Teacher perspectives and experiences of assessment literacy in Victorian junior secondary schools. *Australian Journal of Education*, 68(1), 5–22. <https://doi.org/10.1177/00049441231214022>
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3–17. <https://doi.org/10.1080/09695940701876003>
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97–111. <https://doi.org/10.1016/j.learninstruc.2008.02.003>
- Brown, G. T. L., & Remesal, A. (2017). Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers. *Studies in Educational Evaluation*, 55, 68–74. <https://doi.org/10.1016/j.stueduc.2017.07.003>
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (1st ed., pp. 3–17). Routledge. <https://doi.org/10.4324/9780203874851>
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149–166. <https://doi.org/10.1080/0969594X.2014.898128>
- DeLuca, C., Holden, M., & Rickey, N. (2025). From challenge to innovation: A grassroots study of teachers' classroom assessment innovations. *British Educational Research Journal*, 51(1), 93–114. <https://doi.org/10.1002/berj.4065>
- Fernández-Ruiz, J., & Panadero, E. (2020). Comparison between conceptions and assessment practices among secondary education teachers: More differences than similarities (Comparación entre concepciones y prácticas de evaluación en profesores de Educación Secundaria: más diferencias que semejanzas). *Journal for the Study of Education and Development: Infancia y Aprendizaje*, 43(2), 309–346. <https://doi.org/10.1080/02103702.2020.1722414>
- Fraser, B. J. (2012T). Classroom learning environments: Retrospect, context and prospect. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 1191–1239). Springer. https://doi.org/10.1007/978-1-4020-9041-7_79
- Garazo, M., & Alonso-Tapia, J. (2016). *Perfil motivacional del profesorado en relación con los alumnos: desarrollo de instrumentos de evaluación (Teachers' motivational profile in relation to students: Development of assessment instruments)* [Unpublished master's thesis]. Universidad Autónoma de Madrid.
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169–207. <https://doi.org/10.1080/0969594032000121270>
- Harris, L. R., & Brown, G. T. L. (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. <https://doi.org/10.1080/09695940903319745>
- Hay, P., & Penney, D. (2013). *Assessment in physical education: A sociocultural perspective*. Routledge. <https://doi.org/10.4324/9780203133163>
- IBM Corp.. (2021). *IBM SPSS Statistics for Windows (Version 28.0)*. IBM Corp.
- Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001>
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255–276. <https://doi.org/10.1080/10476210.2011.593164>
- Looney, J. W. (2011). *Integrating formative and summative assessment: Progress toward a seamless system?* (OECD Education Working Papers, No. 58). OECD Publishing. <https://doi.org/10.1787/5kghx3kbl734-en>
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18(1), 57–76. <https://doi.org/10.1080/13573322.2012.713860>
- McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 2–16). SAGE Publications, Inc. <https://doi.org/10.4135/9781452218649.n1>
- Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Nurmi, J.-E. (2012). Students' characteristics and teacher–child relationships in instruction: A meta-analysis. *Educational Research Review*, 7(3), 177–197. <https://doi.org/10.1016/j.edurev.2012.03.001>
- Panadero, E., Brown, G. T. L., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, 21(4), 365–383. <https://doi.org/10.1080/0969594X.2014.919247>
- Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology*, 99(1), 83–98. <https://doi.org/10.1037/0022-0663.99.1.83>
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, 38(3–4), 84–92. <https://doi.org/10.1016/j.stueduc.2012.06.003>
- Reeve, J. (2006). Teachers as facilitators: What autonomy-supportive teachers do and why their students benefit. *The Elementary School Journal*, 106(3), 225–236. <https://doi.org/10.1086/501484>
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *The Curriculum Journal*, 18(1), 27–38. <https://doi.org/10.1080/09585170701292133>
- Remesal, A., & Brown, G. T. L. (2015). Conceptions of assessment when the teaching context and learner population matter: Compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, 30, 331–347. <https://doi.org/10.1007/s10212-014-0236-3>
- Retelsdorf, J., Butler, R., Streblow, L., & Schiefele, U. (2010). Teachers' goal orientations for teaching: Associations with instructional practices, interest in teaching, and burnout. *Learning and Instruction*, 20(1), 30–46. <https://doi.org/10.1016/j.learninstruc.2009.01.001>
- Rubie-Davies, C. M., Flint, A., & McDonald, L. G. (2012). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology*, 82(2), 270–288. <https://doi.org/10.1111/j.2044-8279.2011.02025.x>
- Snijders, T. A. B., & Bosker, R. J. (2012F). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE.

- Spilt, J. L., Koomen, H. M. Y., & Thijs, J. T. (2011). Teacher wellbeing: The importance of teacher–student relationships. *Educational Psychology Review*, 23(4), 457–477. <https://doi.org/10.1007/s10648-011-9170-y>
- 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. <https://doi.org/10.1080/02602930500099102>
- William, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, 99(1), 181–193. <https://doi.org/10.1037/0022-0663.99.1.181>