



OPEN ACCESS

EDITED BY

Kuan-Yu Jin,
Hong Kong Examinations and
Assessment Authority,
Hong Kong SAR, China

REVIEWED BY

Jordi Colomer,
University of Girona, Spain
Ching-Lin Shih,
National Sun Yat-sen University, Taiwan

*CORRESPONDENCE

Eric Jeisy
✉ eric.jeisy@phsg.ch

RECEIVED 21 November 2025

REVISED 13 February 2026

ACCEPTED 16 February 2026

PUBLISHED 03 March 2026

CITATION

Jeisy E, Berthold C and
Baumgartner M (2026) Pre-service
teachers' situation-specific skills
regarding classroom management in
physical education: validation of a
video-based test instrument.
Front. Educ. 11:1751591.
doi: 10.3389/feduc.2026.1751591

COPYRIGHT

© 2026 Jeisy, Berthold and
Baumgartner. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Pre-service teachers' situation-specific skills regarding classroom management in physical education: validation of a video-based test instrument

Eric Jeisy*, Clemens Berthold and Matthias Baumgartner

Institute of Physical Education, Sports and Health, St.Gallen University of Teacher Education, St. Gallen, Switzerland

Introduction: Situation-specific skills, comprising Perception, Interpretation, and Decision-making (PID), are considered a central facet of teachers' professional competence, linking professional knowledge to observable performance. However, empirically validated instruments to assess these skills in the specific context of Classroom Management (CM) in Physical Education (PE) are lacking. This study details the development and validation of a video-based test instrument designed to measure pre-service PE teachers' CM-related PID.

Methods: The study utilized a sample of $n = 877$ pre-service PE teachers from four Swiss universities of teacher education. The instrument includes 10 unscripted video vignettes with 150 closed-format items. Psychometric properties were analyzed using Item Response Theory (IRT), specifically comparing 1PL, 2PL, and 3PL models, as well as testing for dimensionality and differential item functioning (DIF).

Results: Findings supported a unidimensional 2PL model as the most parsimonious and robust solution, despite theoretical assumptions of a tripartite (P-I-D) structure. The instrument demonstrated acceptable reliability ($EAP = 0.73$) and good local fit. A shortened version (7 vignettes) was successfully validated ($r = 0.81$ with the full version), enhancing test economy. Construct validity was supported by a significant moderate correlation with CM-related professional knowledge ($r = 0.29$). Criterion-related validity was supported, as the test successfully differentiated between students in their first and third years of study.

Discussion: The results indicate that the instrument is a valid and efficient tool for assessing CM-related situation-specific skills in teacher education. The unidimensional structure suggests that closed-format video tests primarily capture a holistic, knowledge-based reasoning ability. The study discusses limitations regarding ceiling effects and provides implications for using the instrument in future intervention studies and diagnostic contexts.

KEYWORDS

classroom management, physical education, professional competence, situation-specific skills, teacher noticing, test development, test validation, video vignettes

1 Introduction

Situation-specific skills of Perceiving, Interpreting, and Decision making (PID) are a core component of teachers' professional competence and are empirically linked to instructional quality and student learning (Blömeke et al., 2022; König and Kramer, 2016; König et al., 2021). Within the competence structure model by Blömeke et al. (2015), and its adaptation for Physical Education (PE; Baumgartner, 2022), PID skills are conceptualized as the situation-specific competence facet linking professional knowledge to observable teaching performance. Figure 1 summarizes this competence continuum and situates the focus of the present study.

Classroom Management (CM) constitutes a key professional demand in PE to establish and maintain a productive learning environment (Cothran and Kulinna, 2015). Although CM is often discussed as a generic dimension of teaching quality, PE poses distinct requirements due to its dynamic and spatially distributed learning setting (e.g., movement, noise, transitions, equipment handling, and safety). These subject-specific conditions increase demands on teachers' monitoring, coordination, and rapid situational decision-making (Baumgartner et al., 2020; Ryan and van der Mars, 2022).

Despite the relevance of CM and situation-specific PID skills for teacher education, validated instruments that assess CM-related PID in a conceptually aligned manner are scarce, and instruments for the subject-specific context of PE are lacking. Moreover, existing research frequently relies on combinations of instruments that are not optimally aligned in terms of conceptual validity (Gold et al., 2024). To address this gap within the project *From Knowledge to Performance in Physical Education* (WiPe-Sport; Baumgartner et al., 2025), the presented CM-related PID test was developed alongside a CM-related knowledge measure (Berthold et al., 2025a) and complements an established observation instrument for CM-related performance (Baumgartner et al., 2020), together forming a coherent assessment framework across the competence continuum (see Figure 1).

Importantly, the present study does not aim to empirically test the full competence structure model or the complete knowledge–PID–performance chain. Rather, it focuses on the psychometric validation of a newly developed video-based test designed to assess pre-service PE teachers' CM-related situation-specific PID skills.

Accordingly, this paper reports on the instrument's reliability, dimensional structure, and construct and criterion validity using Item Response Theory (IRT). After outlining the theoretical background, we describe instrument development and content validation, present the validation analyses, and discuss implications and limitations for assessing CM-related situation-specific skills in PE teacher education.

2 Theoretical framework and state of research

2.1 Teachers' professional competence and teacher education

Teachers' professional competence can be understood as a developmental construct encompassing cognitive and affective dispositions and skills necessary for effective teaching (Baumgartner, 2022; Kaiser and König, 2019; Kaiser et al., 2025). The ongoing debate in educational measurements about how teachers' professional competence can be assessed has, in recent years, been shaped by integrative approaches that seek to combine the analytic measurement of aspects of professional competency with situation-specific skills and behavioral performance indicators (Baumgartner, 2018; Baumgartner, 2022; Blömeke et al., 2015; Krauss et al., 2020; Seidel et al., 2024).

A widely cited competence structure model in this context is that of Blömeke et al. (2015). This model, along with its adaptation and extension for the field of PE (Baumgartner, 2022), serves as the theoretical framework for the development of the instrument within the WiPe-Sport research project (Baumgartner et al., 2025). The framework comprises three interrelated competence facets: (1) aspects of competency (i.e., dispositions such as professional knowledge; cf. Baumert and Kunter, 2013), (2) the situation-specific skills of PID, and (3) performance. Together, these facets form a continuum of teachers' professional competence. When performance is conceptualized as the target facet, the aspects of competency and situation-specific PID skills are understood to contribute directly to it, whereas the facet of performance, in turn, supports the positive development of students

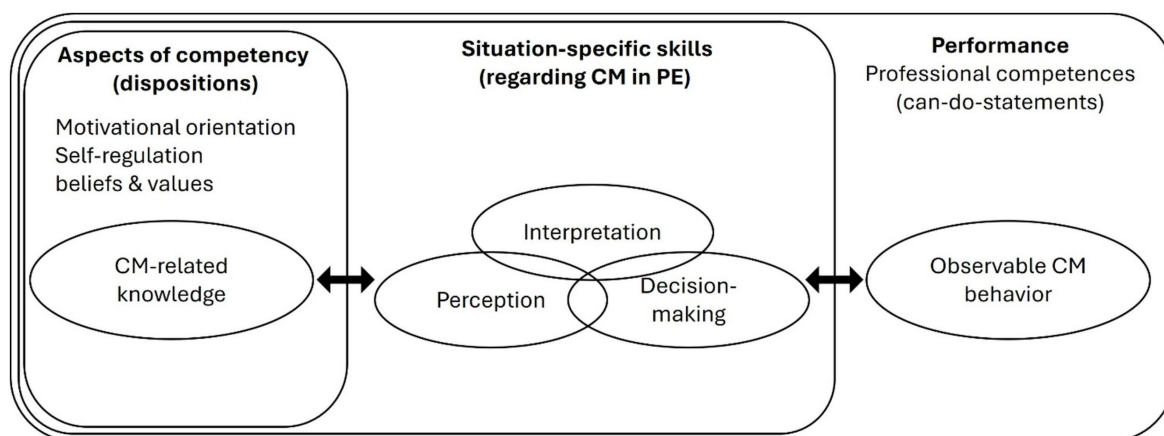


FIGURE 1 Conceptual framework of the WiPe-Sport research project based on the competence structure model (adapted from Blömeke et al., 2015; Baumgartner, 2022), illustrating the mediating role of situation-specific skills (PID) between professional knowledge and performance in the context of PE.

(Baumgartner, 2022). Within this framework, the three facets are additionally connected to distinct competence-areas (e.g., CM, feedback, etc.; Baumgartner, 2022).

Although the different competence facets can be empirically distinguished, they are also interrelated and mutually reinforcing (Bauersfeld et al., 2025a; Dückers et al., 2025; Gold and Holodynski, 2017; König, 2015; König et al., 2025). Aspects of professional competence, such as CM-related knowledge, shape teachers' perception and interpretation of classroom situations and enable top-down information processing (Berthold et al., 2025a; Dückers et al., 2025; Seidel et al., 2024). This leads to more efficient information search and improved sense-making (Wolff et al., 2021). Expert teachers, for instance, rely on highly interconnected and flexible cognitive schemata and scripts for CM (Wolff et al., 2021), which supports situation-specific PID skills to support student learning (Gold and Holodynski, 2017; Stahnke and Blömeke, 2021; Wolff et al., 2016, 2017). Research on teacher expertise has consistently shown that novice and expert teachers differ substantially in their professional vision¹ (e.g., Gold and Holodynski, 2017; Reuter, 2017).

The development of teachers' professional competence is a central goal of teacher education (Kaiser and König, 2019). With regard to situation-specific PID skills, it is assumed that this competence facet evolves over time through a combination of (a) increasing domain- and competence area-specific knowledge, (b) more learning opportunities through authentic teaching experiences, and (c) (guided) reflection on one's own or others' teaching practice (e.g., through video analysis) at the interface of theory and practice (Barenthien et al., 2023; Bauersfeld et al., 2025b; Eßling et al., 2023; Junker et al., 2025; Kaiser et al., 2025; Stahnke and Blömeke, 2021; Stahnke et al., 2016; Weyers et al., 2024; Wolff et al., 2016, 2017).

In summary, we consider situation-specific PID skills to be a crucial facet of teachers' professional competence that should be explicitly targeted and developed within teacher education. The test instrument presented and validated in this study serves, among other purposes, to diagnose pre-service PE teachers' developmental levels in situation-specific PID skills within the competence-area of CM.

2.2 Classroom management in physical education

The CM in PE refers to the sum of teacher actions aimed at creating and maintaining a productive learning environment that supports students' cognitive, social-emotional, and motor development (Baumgartner et al., 2020; Brophy, 2006; Cothran and Kulinna, 2015; Putra and Yanto, 2025). These actions involve both proactive and reactive strategies to maximize on-task behavior, support instructional flow, and prevent or respond to disruptions (Emmer and Stough, 2001; Doyle, 2006; Oliver et al., 2011).

The CM is considered a crucial competence-area for teachers, as it can significantly improve students' attention, motivation, engagement, and learning (Korpershoek et al., 2016, 2025; Kunter et al., 2007; Putra and Yanto, 2025). Mastering CM also supports teachers' health, as it can reduce stress and burnout (Aloe et al., 2014; König and

Rothland, 2016; McCarthy et al., 2022), promotes teacher well-being (Aldrup et al., 2018), and job satisfaction (Dicke et al., 2015).

Although CM is a generic aspect of teaching and teaching quality (Praetorius et al., 2018), it poses distinct challenges in PE due to the subject's dynamic, noisy, and physically active environment (Baumgartner et al., 2020; Cothran and Kulinna, 2015; Herrmann and Gerlach, 2020; Ryan and van der Mars, 2022). Teaching in spaces such as gymnasiums, outdoor fields, or swimming pools requires specialized rules, routines, and strategies to manage transitions, coordinate equipment, and ensure student safety (Chappell, 2020; Lawrence, 2020). Active supervision is crucial, requiring PE teachers to move strategically, maintain visibility, and engage frequently with students to manage behavior and maintain rapport (Arbogast and Chandler, 2005; Schuldheisz and van der Mars, 2001; van der Mars et al., 1994). PE teachers must also address the needs of non-participating students and enforce safety protocols appropriate to the physical demands of specific sports (Chappell, 2020).

To systematically account for these specific subject-matter requirements, we adopt the framework by Baumgartner et al. (2020). In this model, CM in PE is conceptualized as a multidimensional construct that explicitly integrates general pedagogical strategies adapted to the specific demands of PE (monitoring, clarity of announcements, smooth transitions, overlapping, group mobilization, momentum, and dealing with disturbances) alongside two distinct PE-specific dimensions: ensuring safety and managing equipment.

2.3 Situation-specific PID skills regarding classroom management

In the competence structure model proposed by Blömeke et al. (2015), the situation-specific PID skills refer to three distinct cognitive processes required to address professional demands of teaching in classroom situations. Applied here within both a competence-area specific (CM) and subject-specific (PE) context, these skills captures (pre-service) PE teachers' ability to identify and *perceive* CM-relevant situations in PE, *interpret* them based on their CM-related knowledge, experience, and beliefs, and *make decisions* on whether to continue the current action or implement alternative courses of action to support student learning (Blömeke et al., 2015, 2022; Kaiser et al., 2015).

In developing the test instrument, we follow this threefold cognitive structure and therefore use the term *CM-related PID* to denote competence-area specific situation-specific PID skills. Alternative frameworks describe similar constructs under different labels, such as *professional vision* (e.g., Gold and Holodynski, 2017; Reuter, 2017) or *teacher noticing* (e.g., König et al., 2022; Langer et al., 2023). Although these approaches often refer to the same foundational work (e.g., Goodwin, 1994; van Es and Sherin, 2002), they differ in scope and emphasis—for instance, in whether *decision-making* is included as a cognitive process, or whether the focus is placed on *perception* (noticing) and *interpretation* (knowledge-based reasoning; see Amador et al., 2021; König et al., 2022; Schicklinski and Reuter, 2023; Sherin et al., 2024; Stahnke et al., 2016).

Based on the systematic literature review by König et al. (2022), our understanding of CM-related PID can be classified within a *cognitive-psychological perspective*. Accordingly, in developing our test instrument, we specifically address the measurement of the cognitive processes of pre-service PE teachers' situation-specific skills. This theoretical focus can, inter alia, be distinguished from a *socio-cultural perspective*, which emphasizes the social and situated nature of teacher

¹ The term *professional vision* is used here as adopted by the respective authors. Conceptual and terminological distinctions between different approaches to assessing situation-specific skills are discussed in section 2.3.

noticing (Goodwin, 1994; Sherin and van Es, 2008) and from an *expertise-related perspective* focusing on differences between experts and novices (Reuker, 2017; Roose et al., 2018).

Research has shown that CM-related PID represents a learnable competence facet that can be developed during teacher education (Bauersfeld et al., 2025b; Gold et al., 2013; Hellermann et al., 2015; Junker et al., 2025; Kramer et al., 2017; Orschulik, 2021; Weber et al., 2020; Weber et al., 2018) and expert teachers tend to have better CM-related PID than novice teachers (Stahnke and Blömeke, 2021; Wolff et al., 2016, 2017).

Studies grounded in the competence-as-a-continuum framework (Blömeke et al., 2015) indicate that CM-related knowledge is significantly linked to CM-related PID (Gippert et al., 2022; Gold and Holodynski, 2017; König, 2015; Weyers et al., 2024). This relationship is complex and depends on the type of knowledge assessed (declarative vs. procedural), the specific cognitive process within PID, and the level of professional development (pre-service vs. in-service; Junker et al., 2021; Weber et al., 2023). Declarative knowledge, for instance, supports the perception of relevant classroom events, whereas integrated conceptual knowledge appears essential for interpretation and decision-making (Junker et al., 2021). Moreover, there is evidence that CM-related PID predicts various dimensions of teaching quality and student learning outcomes (Blömeke et al., 2022; König and Kramer, 2016; König et al., 2021). However, the empirical evidence in this regard remains limited and inconsistent, as some studies have found no or only partial significant relationship between CM-related PID and performance (Baumgartner et al., 2025; Gold et al., 2021; Junker et al., 2021).

Overall, it can be concluded that situation-specific PID skills represent a theoretically and conceptually multifaceted construct that is highly relevant for teachers' professional development in general and for CM competence in particular.

2.4 Measurement of CM-related situation-specific skills

There are several validated instruments to measure situation-specific skills in a standardized way, ranging from *analytical* (e.g., van Es and Sherin, 2002) or *holistic* testing methods (e.g., Roose et al., 2018) to data collection approaches *on action* (e.g., video-based testing instruments, Gold and Holodynski, 2017) and *in action* (e.g., eye-tracking studies, van Driel et al., 2022). Among these, video-based instruments have become the most widely used (Gold et al., 2024; Weyers et al., 2023). They typically employ video clips of teaching practice as stimuli for closed- or open-ended items or writing prompts.

A broad range of instruments have been developed to assess CM-related PID employing diverse methodological approaches (Bauersfeld et al., 2025a; Gippert et al., 2022; Gold and Holodynski, 2017; Jamil et al., 2015; König, 2015; Seidel and Stürmer, 2014; Weber et al., 2020; Weyers et al., 2024). The spectrum ranges from closed-test formats (e.g., Gold and Holodynski, 2017) to mixed-method approaches (e.g., Weber et al., 2020) and purely qualitative instruments (e.g., Gippert et al., 2022). Analytical approaches further differentiate situation-specific skills into two (Gold and Holodynski, 2017), three (Weber et al., 2020), or four (Gippert et al., 2022) cognitive processes, each empirically assessed as a separate subfacet. One instrument focuses exclusively on a single subfacet (decision-making; Weyers et al., 2024), supplementing an

existing instrument (CME-PI; König, 2015). While some instruments focus solely on CM-related PID, others simultaneously capture additional teaching aspects such as instructional support (Bauersfeld et al., 2025a). In the field of PE, empirically validated instruments to measure teacher noticing are available (e.g., Langer et al., 2023, 2024), but no one specifically addresses CM.

Review studies highlight substantial differences in the development and design of video-based measures (Gold et al., 2024; Weyers et al., 2023). Overall, instruments using open-ended prompts, multimodal foci, or extended video sequences are theoretically better suited to capture complex cognitive processes such as perception (noticing) and decision making (Bauersfeld et al., 2025a; Dückers et al., 2022; Gold et al., 2024; Weber et al., 2020; Seidel and Stürmer, 2014). In closed test formats guided attention through rating items simplify the perception and selection of relevant classroom events and decision-making is typically measured only in terms of evaluating potential situational action alternatives (cf. Weber et al., 2020 for a critique). Consequently, such highly standardized approaches tend to be more strongly aligned with the cognitive process of interpretation or knowledge-based reasoning (cf. Gold et al., 2024; Weyers et al., 2023).² Nevertheless, highly structured, closed-task approaches offer advantages in construct alignment, scoring objectivity, and test efficiency (Bauersfeld et al., 2025a; Eßling et al., 2024; Seidel and Stürmer, 2014; Weyers et al., 2023).

3 Development and validation of the CM-related PID test instrument

For reasons of test efficiency and construct validity within the WiPE-Sport research project, the CM-related PID test and the CM-related knowledge test were developed as highly standardized instruments with close reference to the K1a-Pe-Sport observation instrument (Baumgartner et al., 2020), which is employed in the project to assess pre-service PE teachers' CM-related performance (Baumgartner et al., 2025). While the knowledge test in paper-and-pencil format assesses non-situated declarative knowledge about effective CM practices (Berthold et al., 2025a), the video-based PID test is expected to capture more proximal situation-specific skills, drawing primarily on procedural knowledge (Berthold et al., 2025a; Seidel et al., 2024; Weyers et al., 2023).

Before presenting the objectives and research questions of the empirical validation study, we first introduce the instrument and outline its development. The content validation procedure, particularly regarding the derivation of PE-specific characteristics, has been described in another publication (Baumgartner et al., 2023). In the present paper, we summarize this process and its key results in the next sections for a better understanding.

3.1 Instrument development and description

From the data pool of 60 videotaped primary school PE lessons, a total of 10 unscripted video vignettes (duration

² Contrary to this assumption, Eßling et al. (2024) showed that closed test formats tend to evoke recall rather than interpretative cognitive processes, requiring selective perception within the situation (Seidel and Stürmer, 2014).

TABLE 1 Video-based instrument for measuring CM-related situation specific PID abilities in PE teaching.

Vignette	Duration (min:s)	Subject specific content (topic)	CM dimensions ^a	Items (n)
1	01:19	Gymnastics (Inverted hang on rings)	1,2	16
2 ^b	01:34	<i>Dance (Group choreography)</i>	3,7	16
3	01:35	Gymnastics (Cartwheel)	9,5	13
4	02:15	Games (Tag game)	4,1	13
5	01:33	<i>Circus (Balancing tasks)</i>	8,6,1	15
6	02:01	Games (Throwing and catching)	3,7,8	14
7	03:27	Games (Tag game)	1,6	11
8	01:31	<i>Gymnastics (Body tension & awareness)</i>	4,9	15
9	01:44	Gymnastics (Swinging on rings)	5,2,3	16
10	02:54	Games (Tag game)	9,8,6,1,7	21

^aMonitoring (1); Ensuring safety (2); Clarity of announcements (3); Dealing with disturbances (4); Managing equipment (5); Momentum (6); Smooth transitions (7); Overlapping (8); Group mobilization (9).

^bVideo vignettes presented in italics (2, 5, 8) appear exclusively in the full-length version of the test instrument.

1:19–3:27 min.) were developed. The videos were recorded using both static and mobile camera perspectives (Jeisy, 2011). The teaching sequences were selected to ensure that all nine dimensions of good CM in PE (e.g., monitoring, clarity of instructions, and ensuring safety) were represented at least twice and that a broad range of subject-specific content areas (gymnastics, dance, circus, games) was included to enhance ecological validity (see Table 1; Baumgartner et al., 2023). The three cognitive processes of PID are each assessed using dichotomous items. Participants' responses were coded as correct or incorrect by comparing them with an expert-defined master rating (see Section 3.2).

3.1.1 Measurement of perception

To record the CM-related PID subfacet perception, after watching a video vignette, participants are asked to indicate which CM dimensions they perceive as relevant in the observed teaching sequence. An introductory text emphasized that the teaching sequences may include both effective and improvable CM behaviors. Additionally, short definitions of CM are provided for each video vignette to ensure that the research concepts are being assessed (cf. Gippert et al., 2022). "What aspects of classroom management did you find especially noteworthy in the teaching sequence you observed?" was the prompt for participants. Selected CM dimensions (1–6) in PE were presented, and participants had to judge whether each represented a correct or incorrect response option (e.g., monitoring, ensuring safety; see Figure 2).

3.1.2 Measurement of interpretation

For the subfacet interpretation, participants must draw conclusions from observations of the relevant CM-related dimension in each teaching sequence (cf. Hellermann et al., 2015). To address this, two to three items were developed for each CM dimension represented in the video. After analyzing the teaching sequence, participants assessed whether the described CM actions were correct: "Which of the following CM-related teacher actions occurred?" (Example item for monitoring: "The teacher focuses on the group doing gymnastics on the rings while keeping an eye on the class"; see Figure 3).

3.1.3 Measurement of decision-making

Decision-making is assessed by evaluating possible teaching actions, following Blömeke et al. (2016). Participants are asked, "If you were the teacher in this situation, what would you do differently to improve CM?" For each CM-related dimension, two or three action alternatives are provided, and participants are asked to judge whether each is appropriate for the given situation (example item for monitoring: "When instructing and demonstrating, I position myself to keep an eye on the small group playing"; see Figure 4).

3.2 Content validation

Content validation was established through a three-phase Delphi study (Häder, 2014) with eight CM-in-PE experts and a pretest with 54 pre-service PE teachers.³ Expert feedback informed item revisions and led to the establishment of a consensual master rating, while the pretest confirmed usability and item clarity but indicated low item difficulty and substantial time demands. Since all video vignettes proved to be effective in terms of content validity, none were excluded at this stage of test development. The video-based test instrument for the validation study comprised 10 video vignettes and 161 items, balanced across the three cognitive demands of PID and the nine CM dimensions (Baumgartner et al., 2023). Throughout the Delphi process, experts specifically evaluated whether vignettes and items adequately reflected PE-typical CM challenges (e.g., spatial dispersion, transitions, equipment and safety), thereby strengthening the instrument's subject-specific content validity.

3.3 Study objectives, and research questions

The present study aims to examine the internal psychometric properties, reliability, and dimensional structure of the test instrument, as well as its construct and criterion validity (Gold et al., 2024).

³ Pre-service PE teachers refer to student teachers who are being prepared to teach PE at the kindergarten and primary school levels as one subject among others, and who are therefore not specialized PE subject teachers.

What aspects of classroom management did you find especially noteworthy in the teaching sequence you observed?



- Safety
- Smooth transitions
- Monitoring
- Overlapping
- Dealing with disruptions
- Use of material

FIGURE 2

Screenshot of a video vignette representing monitoring and ensuring safety as teaching-relevant CM dimensions with the question and items to measure pre-service PE teachers' perception skills.

Furthermore, the study investigates whether the instrument can be shortened by reducing the number of video vignettes, to facilitate smoother implementation within the WiPe-Sport intervention study (Baumgartner et al., 2025).

RQ1: Does the instrument provide a reliable measure of CM-related PID in PE (1a), and can the three cognitive demands of PID be empirically distinguished (1b)?

We expect the test instrument to demonstrate sufficient reliability when measuring CM-relevant PID, which is a prerequisite for its use in an intervention study (Weyers et al., 2024). In line with previous research (Bastian et al., 2024; Dückers et al., 2022; Gippert et al., 2022; Junker et al., 2021, 2025), we anticipate that the cognitive processes assessed by the CM-related PID test can be empirically differentiated into distinct subfacets.

RQ2: Can the video-based test instrument be shortened by reducing the number of video vignettes or items, without compromising its reliability or IRT-based measurement properties?

To enhance test efficiency, we explore potential reductions in the number of vignettes or items while maintaining representation of all nine CM dimensions and ensuring comparable results in the measurement of individual skills.

RQ3: Is there a correlation between CM-related knowledge and CM-related PID test scores?

Based on the theoretical framework of knowledge-based situation-specific PID skills and previous empirical findings (cf. Müller and Gold, 2023), we hypothesize a significant positive correlation between the two competence facets. Demonstrating such a correlation would

provide evidence for the test's construct validity as CM-relevant knowledge and CM-relevant PID are not the same construct but closely related (e.g., Junker et al., 2025; Weyers et al., 2024).

RQ4: Do pre-service PE teachers' CM-related PID differ significantly among students in their first, second, and third years of study?

In Switzerland, teacher education follows a one-phase model, meaning that theoretical coursework (e.g., on CM) alternates with internships and practice-oriented courses that closely link theory and practice (see Section 2.1). Therefore, we expect third-year students to outperform first-year students on the CM-relevant PID test. Demonstrating such differences would provide evidence for the test's criterion validity (Gold et al., 2024; Franz et al., 2019).

4 Method

4.1 Participants

The pilot study included 877 pre-service PE teachers specializing in kindergarten and primary education (740 female, 130 male, 7 divers), with a mean age of 23.4 years ($SD = 4.0$, range = 18–54). Participants were recruited from four Swiss Universities of Teacher Education (UTES); St. Gallen UTE [$n = 473$], Lucerne UTE [$n = 357$], Fribourg UTE [$n = 41$], Grisons UTE [$n = 6$]. They were in their first ($n = 277$), second ($n = 283$), or third ($n = 313$) year of study (four not reported). All participants were enrolled in bachelor-level teacher education programs qualifying them to teach PE at the kindergarten or primary level. Although the participants attended general seminars on pedagogy and PE as part of their

Which of the following CM-related teacher actions occurred?



- The teacher positions herself/himself in the room to have a good overview of what is happening in the class.
- By dividing the class into two groups, the teacher's presence is increased for the small group working on the rings.
- The teacher focuses on the group doing gymnastics on the rings while keeping an eye on the class.
- The equipment and materials used are selected in such a way as to ensure the safety of the students.

FIGURE 3 Screenshot of a video vignette representing monitoring and ensuring safety as teaching-relevant CM dimensions with the question and items to measure pre-service PE teachers' interpretation skills.

If you were the teacher in this situation, what would you do differently to improve CM? Select the strategies you think would be most effective.



- I will forgo instructing and demonstrating the correct grip for ring gymnastics in order to devote more time to properly instructing the spotter's grip.
- When instructing and demonstrating, I position myself to keep an eye on the small group playing.
- Instead of thin mats, I will place thicker mats under the rings. This will automatically increase safety for the children doing gymnastics.
- When demonstrating the spotter's position and grip, I will have the student perform the required movement so that I can better demonstrate the appropriate

FIGURE 4 Screenshot of a video vignette representing monitoring and ensuring safety as teaching-relevant CM dimensions with the question and items to measure pre-service PE teachers' decisions-making skills.

standard curriculum, they did not receive any specific intervention or preparatory training regarding this video-based instrument prior to data collection. The test therefore serves as a diagnostic

tool to assess the status quo of their skills within their regular education.

4.2 Measurement

4.2.1 CM-related PID

To assess CM-related PID in PE, we used the previously described instrument consisting of 10 video vignettes and 150 dichotomous items addressing the three cognitive demands of P-I-D. Following IRT analysis, we also created a shortened version of the test instrument (7 video vignettes, 104 Items). Both tests, the short and the extensive versions were used to assess construct and criterion validity.

4.2.2 CM-related knowledge

The CM-related knowledge test, developed in parallel with the CM-related PID test, was also validated in this pilot study. It measures pre-service PE teachers' declarative, non-situated knowledge of effective CM practices, focusing on the nine dimensions outlined by Baumgartner et al. (2020). This knowledge was assessed using dichotomous items. The test comprises 79 items, and its content validity was also ensured through a Delphi study and a pretest (Baumgartner et al., 2023). An example distractor item related to the monitoring dimension is: "To monitor the classroom, a teacher should choose a fixed position to see all students." Berthold et al. (2025a) reported psychometric evidence supporting a unidimensional construct within a two-parameter item response theory (2PL-IRT) framework, with reliability indices indicating adequate measurement precision ($EAP = 0.603$; $WLE = 0.569$).

4.3 Data collection

Data was collected online between March and April 2022 using a questionnaire that combined the CM-related PID and CM-related knowledge tests. A third section gathered personal information, including participants' teaching experience in PE and their self-assessed CM-related competence. The survey was administered via the LimeSurvey platform (LimeSurvey GmbH, n.d.).

To balance participant burden and statistical power for IRT-based analyses, items were randomly assigned. For the PID test, participants were randomly assigned to complete four of the 10 video vignettes (58–71 items), resulting in 316–374 responses per vignette. To control for potential order effects, the sequence of the two competence facets (PID test before the knowledge test and vice versa) was randomized. On average, participants required 33 min ($SD = 13$) to complete the entire survey, spending approximately 17.1 min ($SD = 8.0$) on the PID test and 5.4 min ($SD = 4.5$) on the knowledge test.

A subset of the participants completed the assessment as part of a university course (St. Gallen UTE, Lucerne UTE), whereas others participated independently in their free time (Fribourg UTE, Grisons UTE). In total, 1,473 pre-service PE teachers registered for the study, of whom 1,076 submitted complete responses, resulting in a response rate of 73%. To ensure data quality, only cases in which all items were completed, and the total processing time fell within a predefined realistic range were included. These thresholds were based on empirical estimates of video duration and task completion time, corresponding to 15 min overall – at 2 min for the knowledge test and 8 min for the PID test. After data screening and cleaning procedures, 877 participants (59%) were retained for analysis.

4.4 Data analysis

All items were dichotomously scored (0 = incorrect, 1 = correct) by comparing participants' responses with the master rating validated previously (Baumgartner et al., 2023). Items with extremely high or low solving rates and high expert ambiguity during pretest were re-evaluated prior to analysis.

Psychometric analyses followed an IRT-based approach to examine internal structure and reliability of the PID test. Model selection followed a confirmatory approach, comparing different *a priori* specified, theoretically and design-informed model structures. Validity examination was complemented by inferential statistics (RQ1; Moosbrugger and Kelava, 2020).

The most appropriate IRT model was identified by comparing: (a) models with different parameter specifications, (b) potential multidimensional structures and (c) local and differential item fit. (a) The three-parameter logistic (3PL) model was compared with the simpler, nested two- (2PL) and one-parameter (1PL) models. The 3PL was theoretically expected to fit best, given the dichotomous items based on diverse video vignettes. (b) Using the most parsimonious model, we then examined whether more complex multidimensional structures improved fit. The one-dimensional solution was compared with (1) a theoretically grounded three-dimensional model reflecting the subfacets of PID, (2) a nine-dimensional model representing the CM-related content dimensions, and (3) a ten-dimensional testlet model grouping items by video vignettes (cf. Gold and Holodynski, 2017). Model comparisons relied on likelihood ratio (LR) tests, Akaike (*AIC*) and Bayesian (*BIC*) information criteria, expected a posteriori (*EAP*) and weighted likelihood (*WLE*) reliability estimates. (c) Local item fit was evaluated using weighted likelihood estimates as well as infit and outfit statistics, ensuring values remained within acceptable ranges (0.8–1.2, or *t*-standardized values between -1.96 and 1.96 ; Bond et al., 2021). Items showing substantial misfit were examined individually. Because ecological validity was prioritized, items representing essential aspects of a vignette were retained even when discrimination values were low, acknowledging that the 2PL model weights them accordingly. Differential Item Functioning (DIF) analysis using the Mantel–Haenszel test was conducted to examine whether item responses were conditionally independent of demographic variables such as gender, mother tongue, university placement, and participation type (Diaz et al., 2021).

An Item-Person-Map (Wright Map) was employed to examine how well the distribution of item difficulty matched the range of participant ability levels, thereby assessing the overall adequacy of the test for the given sample. On this map, the average person's ability is positioned at zero on the logit scale, with item difficulties displayed in relation to this reference point. When a person's ability level coincides with an item's position on the scale, the probability of answering that item correctly is approximately 50%. This graphical representation enables an intuitive understanding of how item difficulty aligns with participant ability (cf. Bond et al., 2021).

Finally, a shorter version of the instrument was created by reducing the number of video vignettes while maintaining representation of all 9 CM dimensions (each represented at least two times in the remaining vignettes). IRT models for both the full and the shortened versions were estimated and evaluated in terms of model fit, item parameters, person estimates, and reliability to ensure that the shorter version maintained adequate measurement precision (RQ2).

Construct validity (RQ3) was examined through correlations between person ability estimates from the knowledge and PID tests (long and short version). Criterion-related validity (RQ4) was then assessed by analyzing both test versions abilities to distinguish pre-service PE teachers across study years using ANOVA with post-hoc comparisons.

All analyses were conducted in R Project for Statistical Computing (RRID: SCR_001905) using RStudio (2023.06.01). Model comparisons among 1PL, 2PL, and 3PL specifications were conducted using the mirt package. Analyses of dimensionality and subsequent IRT analyses were performed using the TAM package (Version 4.1.4) with marginal maximum likelihood estimation (e.g., tam.mml.2pl). For multidimensional models, numerical integration used a Monte Carlo approach with 2,000 nodes and 100 iterations (convergence = 1e-4) to ensure stable parameter estimation.

The anonymized data and instrument materials are openly available at SWISS Ubase (Berthold et al., 2025b).

5 Results

Prior to statistical analysis, 11 items were discarded. Two were very easy for participants (solving rate 95 and 97%) and were considered redundant. Two items with a low solving rate (solving rate 21 and 35%) and expert agreement in the content validation was not 100%. In addition, seven items showing high expert ambiguity in the content validation were discarded, as they were initially included for exploratory purposes.

5.1 Model selection and dimensionality

First, we conducted model selection to determine the most appropriate IRT model. The more complex 3PL model was initially estimated but rejected in favor of the simpler, nested 2PL model (Table 2). Although the LR test indicated a statistically significant improvement in model fit for the more complex 3PL model, information criteria (AIC, BIC) suggested that this improvement was offset by increased model complexity. However, the 1PL yielded insufficient reliability, likely reflecting its restrictive assumption of equal item discrimination. Therefore, the 2PL model was selected as the most parsimonious model.

Second, the test instrument was designed to assess pre-service PE teachers' situation-specific PID skills as facet of their CM-related professional competency, which could lead to multidimensionality. Accordingly, several multidimensional solutions were tested: (1) a

three-dimensional model reflecting the theoretical subfacets of perception, interpretation, and decision-making (PID), (2) a nine-dimensional model representing CM content domains; and (3) a ten-dimensional model accounted for potential testlet effects due to shared video stimuli.

Model comparisons indicated that the three-dimensional model provided a significantly better fit than the unidimensional solution [$\Delta\chi^2(3) = 350.52, p < 0.001$], whereas the nine- and ten-dimensional solutions did not improve fit (Table 3). The three-dimensional solution differentiated between P-I-D while accounting for their intercorrelations. While this model showed superior fit and the three subfacets were moderately correlated ($r = 0.10-0.42$), their subfacet reliabilities were below acceptable levels, particularly for decision-making (EAP = 0.68, 0.60, 0.50; WLE = 0.64, 0.55, 0.37).

Given these results, we used the unidimensional 2PL model for all subsequent analyses because it yielded substantially higher reliability and more stable person estimates than the three-dimensional scoring approach. Accordingly, all results reported below are based on the unidimensional model.

5.2 Local item fit and differential item functioning

Local item fit was examined for the unidimensional model using mean-squared residual-based item-fit statistics (tam.fit). Outfit values ranged between 0.85 and 1.05 (t -standardized = -0.31 to 0.66) and Infit values between 0.99 and 1.01 (t -standardized = -0.19 to 0.19), indicating good local fit.

Differential Item Functioning (DIF) analysis using the Mantel-Haenszel test revealed no indications of DIF regarding gender, mother tongue, university placement, or participation type.

5.3 Item-person fit

The Item-Person Map (Figure 5) displays person abilities (left) and item difficulties (right) on the same latent scale. While both appear normally distributed, the majority of items are located below the mean person ability. This suggests that many items were relatively easy for the sample, limiting the test's ability to distinguish higher-ability respondents.

5.4 Short version

A shorter version of the instrument was created by reducing the number of video vignettes from 10 to 7, while maintaining representation of all 9 CM dimensions (each appearing at least twice). This shortened version showed acceptable reliability for a unidimensional solution

TABLE 2 Model comparison and fit indices for 3PL, 2PL and 1PL item response models.

Model	Log-likelihood	AIC	BIC	LR Test (vs. nested simpler model)	EAP reliability	WLE reliability
3PL	-27076	55052	57201	$\chi^2(150) = 208.46, p = 0.001$	0.772	0.865
2PL	-27180	54960	56393	$\chi^2(149) = 850.70, p < 0.001$	0.723	0.717
1PL	-27605	55513	56234	-	0.439	0.440

Loglike, natural logarithm of the likelihood function; AIC, Akaike information criterion; BIC, Bayesian information criterion; EAP, expected a posteriori; WLE, weighted likelihood estimate; PL, parameter logistic. LR test refers to likelihood ratio tests comparing each model with the next simpler nested model. EAP and WLE measures can be interpreted like Cronbach's alpha values.

TABLE 3 Comparison of multidimensional and unidimensional IRT models.

Model	Loglike	Deviance	Npars	Nobs	AIC	BIC	LR test (vs. unidimensional model)
2PL unidimensional	-28500	57001	314	877	57,629	59,129	—
2PL Three-dimensional (Sub-Facets)	-27004.81	54009.61	303	877	54615.61	56062.90	$\chi^2(3) = 350.52, p < 0.001$
2PL Nine-dimensional (Content Dimensions)	-27751.55	55503.11	336	877	56175.11	57780.02	$\chi^2(36) = -1142.98, p = 1.00$ (n.s.)
2PL Ten-dimensional (Testlet)	-28129.12	56258.25	345	877	56948.25	58596.14	$\chi^2(45) = -1898.11, p = 1.00$ (n.s.)

Model, IRT model specification; PL, parameter logistic; loglike, natural logarithm of the likelihood function; Deviance, $-2 \times \log$ -likelihood; Npars, number of estimated parameters; Nobs, number of participants; AIC, Akaike information criterion; BIC, Bayesian information criterion; LR Test (vs. next simpler), likelihood-ratio test comparing each model to the next simpler nested model.

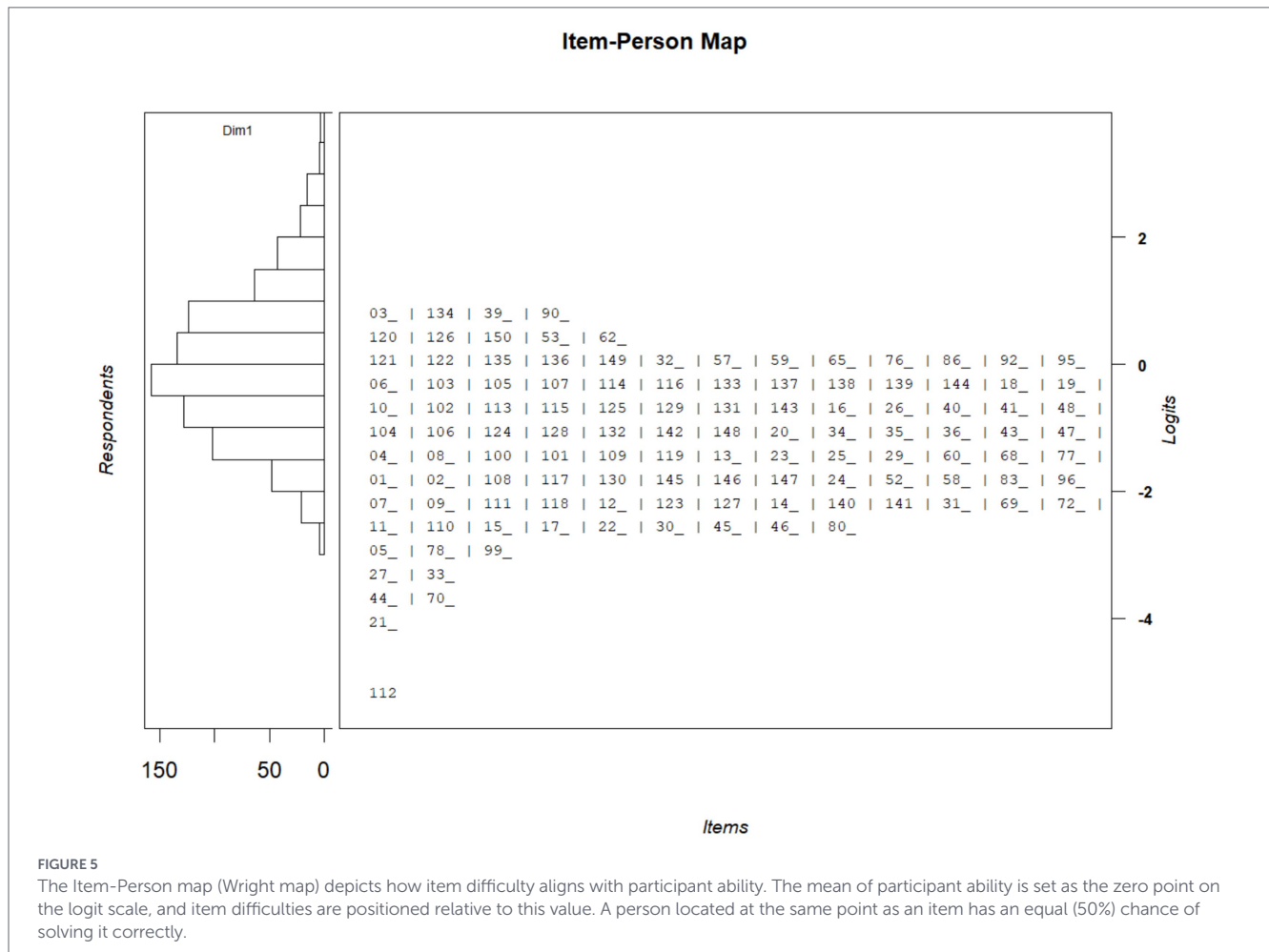


FIGURE 5

The Item-person map (Wright map) depicts how item difficulty aligns with participant ability. The mean of participant ability is set as the zero point on the logit scale, and item difficulties are positioned relative to this value. A person located at the same point as an item has an equal (50%) chance of solving it correctly.

($EAP = 0.674$; $WLE = 0.639$) compared with the full version ($EAP = 0.726$; $WLE = 0.709$). Person estimates from the two versions were strongly correlated ($r = 0.81$; $p < 0.001$), indicating substantial overlap in the underlying ability estimates. Although reliability decreased slightly due to fewer items, the parameter structure and local fit remained stable, indicating that the short form maintained sufficient precision for research applications while improving efficiency.

5.5 Construct validity

Next, we examined construct validity through correlations between person ability estimates from the knowledge and PID tests. For the full PID version, a small but significant positive correlation was found with the knowledge test ($r = 0.29$, $p < 0.001$, 95% CI [0.23, 0.35]). For the short version, the correlation was smaller but still

significant ($r = 0.16, p < 0.001, 95\% \text{ CI } [0.10, 0.23]$). As expected, both PID versions were strongly correlated with each other ($r = 0.81, p < 0.001, 95\% \text{ CI } [0.79, 0.83]$). These results indicate that, while related, the two instruments assess with knowledge and PID two distinct facets of CM competence—supporting the test's construct validity.

5.6 Criterion-related validity (RQ4)

Finally, criterion-related validity was assessed by comparing pre-service PE teachers' test scores across study years. ANOVAs conducted separately for the short and full versions revealed significant effects of study year on test performance, indicating that the instrument distinguishes between different stages of teacher education. For the short version, the effect of study year was significant ($f[2, 859] = 11.94, p < 0.001$). Tukey *post hoc* comparisons showed that second-year students scored higher than first-year students ($\Delta M = 0.33, 95\% \text{ CI } [0.04, 0.63], p = 0.022$), and third-year students scored higher than first-year students ($\Delta M = 0.60, 95\% \text{ CI } [0.31, 0.88], p < 0.001$). The difference between second- and third-year students was not significant ($p = 0.076$).

For the full version, the effect of study year was likewise significant ($f[2, 870] = 7.57, p < 0.001$). Tukey *post hoc* comparisons indicated that third-year students outperformed first-year students ($\Delta M = 0.39, 95\% \text{ CI } [0.15, 0.62], p < 0.001$), whereas differences between first- and second-year ($p = 0.12$) and between second- and third-year students ($p = 0.14$) were not significant.

Together, these results demonstrate that both the full and short versions of the test effectively distinguish between pre-service PE teachers at different stages of their education, providing evidence for criterion-related validity.

6 Discussion

This study validated a video-based test instrument designed to assess CM-related situation-specific PID skills in pre-service PE teachers. Using Item Response Theory (IRT) and multiple validity analyses with a large sample ($n = 877$), the findings provide promising evidence for the instrument's psychometric quality while also highlighting areas for refinement.

The discussion follows the four research questions and integrates them with the broader theoretical context.

6.1 Validation

The analyses (RQ1) showed that a unidimensional two-parameter logistic (2PL) model offered the most parsimonious and psychometrically stable solution for the full instrument. While the three-dimensional model (corresponding to P-I-D) achieved superior fit compared to a strictly unidimensional model, the reliabilities of the subfacets (especially decision-making) were low.

Importantly, retaining a unidimensional solution should not be interpreted as evidence against the theoretical differentiation of perception, interpretation, and decision-making. Rather, the unidimensional score summarizes the shared variance across P-I-D-related item sets and serves as a pragmatic scoring model for the present

instrument. This is also plausible from a cognitive-psychological perspective, as P-I-D draw on partially overlapping knowledge-based resources and are sequentially interdependent in real-time classroom processing (Weyers et al., 2024).

From a measurement perspective, recent methodological work demonstrates that the empirical separability of PID subprocesses is substantially shaped by assessment format (Gold et al., 2024). In the present study, the use of a closed-response format likely contributed to the predominantly unidimensional structure observed in the analyses. Although each video vignette was designed to address perception, interpretation, and decision-making through dedicated questions, the dichotomous response options primarily assessed participants' ability to recall and evaluate CM-relevant classroom events based on their existing knowledge (Eßling et al., 2024). While accurate recall presupposes perceptual awareness (Seidel and Stürmer, 2014), the item format may have shifted the cognitive focus toward knowledge-based reasoning – that is, the situation-specific ability to interpret and judge CM events through a knowledge-driven lens rather than capturing the full spectrum of P-I-D processes (cf. Gold et al., 2024). This aligns with findings that closed, rating-based formats tend to emphasize interpretive and evaluative processes rather than spontaneous perception or *in situ* decision-making (Gold et al., 2024; Weber et al., 2020). Consequently, the instrument may have captured CM-related situation-specific reasoning as an integrated ability more strongly than the distinct cognitive subfacets postulated in the PID framework.

Finally, the Item-Person Map confirms a mismatch between item difficulty and participant ability, as most items fall below the mean person ability. This imbalance indicates that the test may insufficiently capture variation among higher-ability participants, likely due to a ceiling effect (Berthold et al., 2025a). Consequently, the instrument's sensitivity at the upper end of the scale and its validity for assessing advanced levels of CM-related PID appears limited. Interestingly, this high level of demonstrated competence contrasts with the subjective experience often reported in the literature, where pre-service teachers express difficulties with CM (Dicke et al., 2015; Stokking et al., 2003) and criticize insufficient training in teacher education (Stevenson et al., 2020). However, our findings align with other empirical studies that – contrary to self-reports – document relatively high levels of CM-related knowledge (Gippert et al., 2022; Junker et al., 2021; Schlag and Glock, 2019), CM-related PID (Gold et al., 2021; Junker et al., 2021), and even performance among pre-service teachers (Gold et al., 2021). Thus, the observed ceiling effect may not only reflect test ease but also indicate that the target group possesses a more solid foundation in CM skills than their subjective perception suggests.

Beyond these psychometric considerations, the results should be interpreted in light of the subject-specific demands of CM in PE. In PE, CM is closely intertwined with movement dynamics, spatial dispersion, frequent transitions, and equipment logistics, with safety as a central boundary condition. By embedding these PE-typical demands in unscripted PE video vignettes and aligning items with a PE-specific CM framework (including ensuring safety and managing equipment), the present instrument provides a context-sensitive operationalization of CM-related situation-specific PID. This strengthens its applicability for PE teacher education and the WiPe-Sport research project (Baumgartner et al., 2025).

6.2 Test economy

Regarding test economy (RQ2), the shortened version of the instrument (seven video vignettes) demonstrated satisfactory model fit and reliability, with person ability estimates strongly correlating with those from the full version ($r = 0.81$). These findings suggest that the shorter form can serve as a practical alternative in time-sensitive research settings, such as intervention studies within the WiPe-Sport project (Baumgartner et al., 2025).

Nevertheless, some caution is warranted, as slight losses in precision, discriminative power, and subject-specific (PE) ecological validity may occur. Although all 9 CM dimensions remain represented at least twice in the short version, its subject-specific (PE) scope in the video sequences is restricted to the content areas *games* and *gymnastics*. In contrast, the long version also includes *dance* and *circus*, providing a more comprehensive representation of PE teaching diversity (see Table 1).

6.3 Construct validity

As hypothesized (RQ3), correlations between CM-related knowledge and CM-related PID scores were significant yet moderate ($r = 0.29$ for full version; $r = 0.16$ for short version). This supports construct validity: while related, knowledge and situation-specific PID skills are not redundant but represent distinct facets of the competence construct (Baumgartner, 2022; Blömeke et al., 2015). The moderate correlation aligns with prior findings (cf. Müller and Gold, 2023), emphasizing that the transition from theoretical understanding to in-situ application requires cognitive integration rather than a simple transfer of knowledge (Berthold et al., 2025a).

6.4 Criterion validity

The test successfully differentiated pre-service PE teachers across study years, providing evidence for its criterion validity (RQ4). Both the full and short versions showed significant improvements from the first to the second and from the first to the third year, although differences between the second and third year were not consistently significant.

We interpret this finding in two ways. First, it supports previous research indicating that situation-specific PID represents a learnable competence facet of teachers' professional competence that develops within teacher education programs combining practical learning opportunities and theoretical coursework (Barenthien et al., 2023; Bauersfeld et al., 2025b; Eßling et al., 2023; Junker et al., 2025; Kaiser et al., 2025; Stahnke and Blömeke, 2021; Stahnke et al., 2016; Weyers et al., 2024; Wolff et al., 2016, 2017). However, it is important to acknowledge that our data stem from cross-sectional cohort comparisons and thus cannot capture individual developmental trajectories. Furthermore, data were collected at several Swiss UTEs, which, although all follow a one-phase teacher education model, may differ in the design and implementation of their programs. Further analyses would be needed to account for these institutional variations and to compare our findings with studies conducted in other countries, where the development of situation-specific PID skills has been investigated within a two-phase teacher education system (e.g., Germany, Bauersfeld et al., 2025b).

Second, the results are encouraging from a practical perspective. The instrument's ability to distinguish between pre-service PE teachers at different stages of their development of professional competence highlights its diagnostic potential for assessing developmental levels of CM-related PID within teacher education. Such differentiation provides valuable feedback for program evaluation and for the targeted design of learning opportunities that foster CM-related PID.

6.5 Limitations and directions for future research

Several limitations of this study must be acknowledged. First, the CM-related PID instrument was validated with pre-service kindergarten and primary school PE teachers enrolled at four UTEs in Switzerland. Given the differences between PE teacher education programs (cf. MacPhail et al., 2019) and the subject-specific nature of the instrument (PE), the findings cannot be generalized to other school levels, subjects, or teacher education systems. Moreover, the applicability of the instrument to in-service PE teachers remains untested and should be examined in future research to determine its relevance across different career stages.

Further methodological limitations concern the limited reliability when distinguishing between the cognitive subprocesses of PID. Consequently, CM-related PID was modeled as a unidimensional construct, limiting fine-grained interpretation. Future developments might therefore consider more open-ended formats that better capture the nuanced differences between the subprocesses (cf. Weyers et al., 2023). Also, the relatively low item difficulty (high solution rates) may cause ceiling effects in more advanced cohorts, reducing sensitivity to growth.

Additionally, focusing on a single competence-area (CM) does not fully reflect the multidimensional reality of teaching (Praetorius and Charalambous, 2018). Effective teaching in PE simultaneously involves multiple, interrelated generic competence-areas such as instructional support, student engagement, and content representation (Bauersfeld et al., 2025a; Dückers et al., 2022, 2025) and PE-specific competence-areas such as cognitive-motor activation (Kruse et al., 2024). The presented CM-related instrument thus assesses only one limited aspect of pre-service PE teachers' situation-specific PID skills. Similarly, the use of a master rating as a reference standard must be interpreted cautiously – within the complexity of real classroom interactions, the notion of “correct” or “incorrect” teaching behavior is inherently debatable (Gold et al., 2024). The results should therefore be viewed as an alignment with a normatively defined yet contestable benchmark of good teaching rather than an absolute measure of professional competence.

Another limitation concerns the instrument's ecological validity. The video-based test captures teachers' situation specific PID skills *on action* – that is, reflective judgments about observed situations – rather than *in action*, where decisions are made in real time within the dynamic flow of classroom interaction (cf. Witt et al., 2024). Consequently, the test reflects only part of the situational complexity inherent in authentic teaching situations.

Finally, recent evidence suggests that individual cognitive characteristics, such as selective attention, may influence performance in video-based PID assessments (Dücker et al., 2025). While situation-specific PID skills remains a primarily knowledge-based process, such findings indicate that general cognitive abilities might moderate the

extent to which individuals can translate knowledge into situational PID. Accounting for these interindividual differences could improve the precision and fairness of future test designs.

Taken together, these limitations point to several directions for future research. Subsequent studies should (a) test the instrument across broader educational contexts and teacher populations, (b) refine item formats to increase cognitive differentiation and difficulty, and (c) explore ways to assess multiple competence-areas simultaneously. Developing integrated test instruments that encompass several generic and PE-specific dimensions of teaching – such as CM, instructional quality, student support and cognitive-motor activation – would offer a more holistic and ecologically valid assessment of pre-service PE teachers' professional competence.

7 Conclusion and practical implications

This study introduces the first validated, video-based instrument for assessing CM-related, situation-specific PID skills among pre-service PE teachers. Grounded in the competence-as-continuum framework (Baumgartner, 2022; Blömeke et al., 2015) and supported by a content validation process (Baumgartner et al., 2023), the instrument demonstrates promising psychometric properties – including good reliability in both full and short forms, meaningful discrimination across training levels, and acceptable construct validity in relation to CM-related knowledge in PE.

Although the distinct measurement of P-I-D still requires further methodological refinement, and ceiling effects remain a concern, the validated instrument nonetheless fills an important gap in the field. It provides a feasible, theoretically grounded, and diagnostically meaningful tool that can be used by educational researchers and teacher educators alike in intervention studies, longitudinal designs, and diagnostic applications.

From a practical perspective, the video-based format and the efficiency of the shortened version make the instrument particularly suitable for integration into PE teacher education programs and research on teacher learning. Given its strong construct alignment with instruments measuring CM-related knowledge (Berthold et al., 2025a) and CM-related performance (Baumgartner et al., 2020), it represents a valuable addition to the toolbox for examining the developmental progression of pre-service PE teachers' professional competence on their path toward becoming expert teachers (Kaiser et al., 2025).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://doi.org/10.48573/j4bn-xr96>.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements.

Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

EJ: Writing – original draft, Software, Conceptualization, Writing – review & editing, Investigation, Formal analysis, Methodology, Validation, Visualization. CB: Investigation, Writing – review & editing, Software, Writing – original draft, Validation, Conceptualization, Methodology, Data curation, Formal analysis. MB: Writing – original draft, Writing – review & editing, Funding acquisition, Investigation, Resources, Methodology, Conceptualization.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Swiss National Science Foundation (SNSF). The grant 192397 has been awarded to Matthias Baumgartner (St.Gallen University of Teacher Education). For further information see: <https://p3.snf.ch/project-192397>.

Acknowledgments

The authors thank Sabine Reuker for her support regarding the content development of the test.

Conflict of interest

The author(s) declared that this work 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 author(s) declared that Generative AI was used in the creation of this manuscript. The Gemini model 2.5 Flash and ChatGPT version 4 were employed for the linguistic revision of this manuscript to enhance its clarity, coherence, and overall readability. This AI-assisted editing was focused on refining sentence structure, improving phrasing, and ensuring academic rigor, while maintaining the integrity of the original content.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

References

- Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., and Trautwein, U. (2018). Student misbehavior and teacher well-being: testing the mediating role of the teacher-student relationship. *Learn. Instr.* 58, 126–136. doi: 10.1016/j.learninstruc.2018.05.006
- Aloe, A. M., Amo, L. C., and Shanahan, M. E. (2014). Classroom management self-efficacy and burnout: a multivariate meta-analysis. *Educ. Psychol. Rev.* 26, 101–126. doi: 10.1007/s10648-013-9244-0
- Amador, J., Bragelman, J., and Superfine, A. (2021). Prospective teachers' noticing: a literature review of methodological approaches to support and analyze noticing. *Teach. Educ.* 99:103256. doi: 10.1016/j.tate.2020.103256
- Arbogast, G., and Chandler, J. P. (2005). Class management behaviors of effective physical educators. *Strategies* 19, 7–11. doi: 10.1080/08924562.2005.11000381
- Barenthien, J., Fiebranz, A., Todorova, M., Möller, K., and Steffensky, M. (2023). Development of professional vision and pedagogical content knowledge during initial teacher education. *Int. J. Educ. Res.* 119:102186. doi: 10.1016/j.ijer.2023.102186
- Bastian, A., König, J., Weyers, J., Siller, H.-S., and Kaiser, G. (2024). Effects of teaching internships on preservice teachers' noticing in secondary mathematics education. *Front. Educ.* 9:1360315. doi: 10.3389/feduc.2024.1360315
- Bauersfeld, J. L., Bourcave, P., Hahn, H., and Gold, B. (2025a). Development and initial validation steps of a standardized video test assessing professional vision of classroom management and instructional support. *Educ. Sci.* 15:749. doi: 10.3390/educsci15060749
- Bauersfeld, J. L., Gold, B., and Holodyski, M. (2025b). Development of classroom management competencies throughout teacher education: a longitudinal study. *Teach. Dev.* 29, 891–911. doi: 10.1080/13664530.2025.2455113
- Baumert, J., and Kunter, M. (2013). "The COACTIV model of teachers' professional competence" 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 (New York: Springer), 25–48.
- Baumgartner, M. (2018). Performanzfortschritt in der Lehrerausbildung durch die Arbeit an eigenen video- und textbasierten Unterrichtsfallen? Eine Interventionsstudie zur Verbesserung des Feedbacks bei angehenden Sportlehrkräften. *Zeitschrift für Erziehungswissenschaft*. 21, 1135–1155. doi: 10.1007/s11618-018-0829-8
- Baumgartner, M. (2022). Professional competence(s) of physical education teachers: terms, traditions, modelling and perspectives. *Germ. J. Exerc. Sport Res.* 52, 550–557. doi: 10.1007/s12662-022-00840-z
- Baumgartner, M., Berthold, C., and Jeisy, E. (2025). The effectiveness of cumulative interventions on the classroom management-related performance of pre-service teachers in physical education: improvement of knowledge and situation-specific skills alone is not sufficient. *Z. Erzieh.* 28, 1477–1500. doi: 10.1007/s11618-025-01358-4
- Baumgartner, M., Jeisy, E., and Berthold, C. (2023). From knowledge to performance in physical teacher education: a Delphi study and a pretest for the content validation of the test instruments. *Swiss J. Educ. Res.* 45, 151–163. doi: 10.24452/sjer.45.2.6
- Baumgartner, M., Oesterheld, V., and Reuker, S. (2020). Konstruktion und Validierung eines multidimensionalen Beobachtungsinstrumentes zur Erfassung der klassenführungsbezogenen Performanzen von sportunterrichtenden Lehrkräften (KlaPe-Sport). *Germ. J. Exerc. Sport Res.* 50, 511–522. doi: 10.1007/s12662-020-00675-6
- Berthold, C., Jeisy, E., and Baumgartner, M. (2025a). Pre-service teachers' knowledge of evidence-based classroom management practices in physical education: a test validation using item response theory. *Front. Educ.* 10:1570510. doi: 10.3389/feduc.2025.1570510
- Berthold, C., Jeisy, E., Zulian, D., and Baumgartner, M. (2025b). *WiPe-Sport: Pilotierung Instrumentenentwicklung (Version 1.0) [Data set]*: FORS. doi: 10.48573/j4bn-xr96
- Blömeke, S., Busse, A., Kaiser, G., König, J., and Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teach. Teach. Educ.* 56, 35–46. doi: 10.1016/j.tate.2016.02.003
- Blömeke, S., Gustafsson, J.-E., and Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Z. Psychol.* 223, 3–13. doi: 10.1027/2151-2604/a000194
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., and König, J. (2022). Opening up the black box: teacher competence, instructional quality, and students' learning progress. *Learn. Instr.* 79:101600. doi: 10.1016/j.learninstruc.2022.101600
- Bond, T. G., Yan, Z., and Heene, M. (2021). *Applying the Rasch model: fundamental measurement in the human sciences*. 4th Edn. New York: Routledge Taylor & Francis Group.
- Brophy, J. (2006). "History of research on classroom management" in *Handbook of classroom management: Research, practice, and contemporary issues*. eds. C. M. Evertson and C. S. Weinstein (Mahwah, NJ: Lawrence Erlbaum), 17–43.
- Chappell, A. (2020). "Teaching safely and safety in physical education" in *Learning to teach physical education in the secondary school*. eds. S. Capel, J. Cliffe and J. Lawrence. 5th ed. (Mahwah, NY: Routledge), 284–305.
- Cothran, D., and Kulinna, P. (2015). "Classroom management in physical education" in *Handbook of classroom management*. eds. S. Capel, J. Cliffe and J. Lawrence. 5th ed. (London: Routledge), 239–260.
- Diaz, E., Brooks, G., and Johanson, G. (2021). Detecting differential item functioning: item response theory methods versus the mantel-Haenszel procedure. *Int. J. Assess. Tools Educ.* 8, 376–393. doi: 10.21449/ijate.730141
- Dicke, T., Elling, J., Schmeck, A., and Leutner, D. (2015). Reducing reality shock: the effects of classroom management skills training on beginning teachers. *Teach. Teach. Educ.* 48, 1–12. doi: 10.1016/j.tate.2015.01.013
- Doyle, W. (2006). "Ecological approaches to classroom management" in *Handbook of classroom management: Research, practice, and contemporary issues*. eds. C. M. Evertson and C. S. Weinstein (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 97–125.
- Dückers, C., Gold, B., and Holodyski, M. (2025). How professional knowledge and general cognitive abilities relate to pre-service teachers' professional vision. *Think. Skills Creat.* 59:101970. doi: 10.1016/j.tsc.2025.101970
- Dückers, C., Hörter, P., Junker, R., and Holodyski, M. (2022). Professional vision of teaching as a focus-specific or focus-integrated skill: conceptual considerations and video-based assessment. *Teach. Teach. Educ.* 117:103797. doi: 10.1016/j.tate.2022.103797
- Emmer, E. T., and Stough, L. M. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educ. Psychol.* 36, 103–112. doi: 10.1207/S15326985EP3602_5
- Eßling, I., Sunder, C., Steffensky, M., and Meschede, N. (2024). Video-based measurement of professional vision: comparison of open and closed formats in the context of instructional support in elementary science education. *Lehrerbild. Pruefstand* 17, 127–149. doi: 10.62350/LTTK9621
- Eßling, I., Todorova, M., Sunder, C., Steffensky, M., and Meschede, N. (2023). The development of professional vision in pre-service teachers during initial teacher education and its relationship to beliefs about teaching and learning. *Teach. Teach. Educ.* 132:104250. doi: 10.1016/j.tate.2023.104250
- Franz, E.-K., Heyl, V., Wacker, A., and Dörfler, T. (2019). Konstruktvalidierung eines Tests zur Erfassung von adaptiver Handlungskompetenz in heterogenen Gruppen. *J. Educ. Res. Online* 11, 116–146. doi: 10.25656/01:18029
- Gippert, C., Hörter, P., Junker, R., and Holodyski, M. (2022). Professional vision of teaching as a focus-specific or focus-integrated skill – conceptual considerations and video-based assessment. *Teach. Teach. Educ.* 117:103797. doi: 10.1016/j.tate.2022.103797
- Gold, B., Förster, S., and Holodyski, M. (2013). Evaluation eines videobasierten Trainingseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Z. Pädagog. Psychol.* 27, 141–155. doi: 10.1024/1010-0652/a000100
- Gold, B., and Holodyski, M. (2017). Using digital video to measure the professional vision of elementary classroom management: test validation and methodological challenges. *Comput. Educ.* 107, 13–30. doi: 10.1016/j.compedu.2016.12.012
- Gold, B., Junker, R., Wissemann, M., Klassen, C., and Holodyski, M. (2021). Are good observers good classroom managers? The relationship between teachers' professional vision and their students' ratings on classroom management. *Int. J. Educ. Res.* 109:101811. doi: 10.1016/j.ijer.2021.101811
- Gold, B., Müller, M. M., Weber, K. E., and Prilop, C. N. (2024). "Characteristics and validation of video-based standardised instruments measuring professional vision: a closer look at the predictive validity of different task formats" in *Teacher professional vision: theoretical and methodological advances*. eds. A. Gegenfurtner and R. Stahnke (London: Routledge), 89–108. doi: 10.4324/9781003370901-10
- Goodwin, C. (1994). Professional vision. *Am. Anthropol.* 96, 606–633. doi: 10.1525/aa.1994.96.3.02a00100
- Häder, M. (2014). *Delphi-Befragungen. Ein Arbeitsbuch*. Wiesbaden: Springer.
- Hellermann, C., Gold, B., and Holodyski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium: Die Wirkung der Analyse eigener und fremder Unterrichtsvideos auf das strategische Wissen und die professionelle Wahrnehmung. *Z. Entwicklungspsychol. Pädagog. Psychol.* 47, 97–109. doi: 10.1026/0049-8637/a000129
- Herrmann, C., and Gerlach, E. (2020). Unterrichtsqualität im Fach Sport – Ein Überblicksbeitrag zum Forschungsstand in Theorie und Empirie. *Unterrichtswiss.* 48, 361–384. doi: 10.1007/s42010-020-00080-w

- Jamil, F. M., Sabol, T. J., Hamre, B. K., and Pianta, R. C. (2015). Assessing teachers' skills in detecting and identifying effective interactions in the classroom. *Elem. Sch. J.* 115, 407–432. doi: 10.1086/680353
- Jeisy, E. (2011). *Basismodelle des Lernens und Lehrens: Kameraskript und Anleitung zur Datenerhebung*. Magglingen: Swiss Federal Institute of Sport Magglingen.
- Junker, R., Dückers, C., Grewe, O., Henke, I., Hörter, P., Jürgens, M., et al. (2025). Promoting professional vision of pre-service teachers: does the focus of analysis matter? *Teach. Dev.*, 1–23. doi: 10.1080/13664530.2025.2517880
- Junker, R., Gold, B., and Holodynski, M. (2021). Classroom management of pre-service and beginning teachers: from dispositions to performance. *Int. J. Mod. Educ. Stud.* 5, 339–363. doi: 10.51383/ijonmes.2021.137
- Kaiser, G., Busse, A., Hoth, J., König, J., and Blömeke, S. (2015). About the complexities of video-based assessments: theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *Int. J. Sci. Math. Educ.* 13, 369–387. doi: 10.1007/s10763-015-9616-7
- Kaiser, G., and König, J. (2019). Competence measurement in (mathematics) teacher education and beyond: implications for policy. *High. Educ. Policy* 32, 597–615. doi: 10.1057/s41307-019-00139-z
- Kaiser, G., König, J., Krauss, S., Buchholtz, N., and Gruber, H. (2025). Are teacher expertise and teacher competence complementary or contradictory research paradigms? Reflections on a possible integration and first conceptualizations. *ZDM* 57, 1077–1092. doi: 10.1007/s11858-025-01743-7
- König, J. (2015). Measuring classroom management expertise (CME) of teachers: a video-based assessment approach and statistical results. *Cogent Educ.* 2:991178. doi: 10.1080/2331186X.2014.991178
- König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Felske née Nehls, C., Musekamp, F., et al. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educ. Stud. Math.* 107, 189–212. doi: 10.1007/s10649-020-10021-0
- König, J., Kaiser, G., Bastian, A., Weyers, J., Buchholtz, N., and Ross, N. (2025). Teacher noticing in inclusive mathematics education: analyzing its structure and expert-novice differences using a novel video-based test instrument. *ZDM* 57, 1109–1123. doi: 10.1007/s11858-025-01723-x
- König, J., and Kramer, C. (2016). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM* 48, 139–151. doi: 10.1007/s11858-015-0705-4
- König, J., and Rothland, M. (2016). Klassenführungswissen als Ressource der Burnout-Prävention? Zum Nutzen von pädagogisch-psychologischem Wissen im Lehrberuf. *Unterrichtswiss* 44, 425–441.
- König, J., Santagata, R., Scheiner, T., Adleff, A.-K., Yang, X., and Kaiser, G. (2022). Teacher noticing: a systematic literature review of conceptualizations, research designs, and findings on learning to notice. *Educ. Res. Rev.* 36:100453. doi: 10.1016/j.edurev.2022.100453
- Korpershoek, H., de Boer, H., and Mouw, J. M. (2025). An update of the meta-analysis of the effects of classroom management interventions on students' academic, behavioral, social-emotional, and motivational outcomes. *Rev. Educ. Res.* 1–41. doi: 10.3102/00346543251361903
- Korpershoek, H., Harms, T., de Boer, H., van Kwijk, M., and Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Rev. Educ. Res.* 86, 643–680. doi: 10.3102/0034654315626799
- Kramer, C., König, J., Kaiser, G., Ligtvoet, R., and Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situations-spezifische Fähigkeiten angehender Lehrkräfte. *Z. Erziehungswiss.* 20, 137–164. doi: 10.1007/S11618-017-0732-8
- Krauss, S., Bruckmaier, G., Lindl, A., Hibert, S., Binder, K., Steib, N., et al. (2020). Competence as a continuum in the COACTIV study: the “cascade model”. *ZDM* 52, 311–327. doi: 10.1007/s11858-020-01151-z
- Kruse, F., Büchel, S., and Brühwiler, C. (2024). Dimensionality of instructional quality in physical education. Obtaining students' perceptions using bifactor exploratory structural equation modeling and multilevel confirmatory factor analysis. *Front. Psychol.* 15:1370407. doi: 10.3389/fpsyg.2024.1370407
- Kunter, M., Baumert, J., and Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learn. Instr.* 17, 494–509. doi: 10.1016/j.learninstruc.2007.09.002
- Langer, W., Bruns, J., and Erhorn, J. (2023). Development and validation of a video-based test instrument to measure noticing with regard to providing individual learning support in inclusive physical education. *J. Educ. Res. Online* 15, 51–73. doi: 10.31244/jero.2023.01.03
- Langer, W., Bruns, J., and Erhorn, J. (2024). “Development and validation of a test to measure situation-specific skills with regard to providing participation in inclusive physical education” in *Teacher professional vision: Empirical perspectives*. eds. R. Stahnke and A. Gegenfurtner (London: Routledge), 232–248. doi: 10.4324/9781003370604-17
- Lawrence, J. (2020). “Lesson organisation and management” in *Learning to teach physical education in the secondary school*. eds. S. Capel, J. Cliffe and J. Lawrence. 5th ed (London: Routledge), 106–127.
- LimeSurvey GmbH (n.d.). *LimeSurvey: an open source survey tool*. Hamburg: LimeSurvey GmbH.
- MacPhail, A., Tannehill, D., and Avşar, Z. (2019). *European physical education teacher education practices*. Aachen: Meyer & Meyer.
- McCarthy, C. J., Mosley, K. C., and Dillard, J. B. (2022). “Teacher stress and classroom management” in *Handbook of classroom management*. eds. E. J. Sabornie and D. L. Espelage. 3rd ed. (London: Routledge), 189–210.
- Moosbrugger, H., and Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion*. 3rd Edn. Berlin: Springer.
- Müller, M. M., and Gold, B. (2023). Videobasierte Erfassung wissenschaftsbasierter Verarbeitens als Teilprozess der professionellen Unterrichtswahrnehmung – Analyse eines geschlossenen und offenen Verfahrens. *Z. Erziehungswiss.* 26, 7–29. doi: 10.1007/s11618-022-01128-6
- Oliver, R. M., Wehby, J. H., and Reschly, D. J. (2011). Teacher classroom management practices: effects on disruptive or aggressive student behavior. *Campbell Syst. Rev.* 7, 1–55. doi: 10.4073/csr.2011.4
- Orschulik, A. B. (2021). *Entwicklung der professionellen Unterrichtswahrnehmung: Eine Studie zur Entwicklung Studierender in universitären Praxisphasen*. Wiesbaden: Springer.
- Praetorius, A. K., and Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM* 50, 535–553. doi: 10.1007/s11858-018-0946-0
- Praetorius, A. K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM Mathemat. Educ.* 50, 407–426. doi: 10.1007/s11858-018-0918-4
- Putra, E., and Yanto, M. (2025). Classroom management: boosting student success – a meta-analysis review. *Cogent Educ.* 12, 1–14. doi: 10.1080/2331186X.2025.2458630
- Reuker, S. (2017). The noticing of physical education teachers: a comparison of groups with different expertise. *Phys. Educ. Sport Pedagog.* 22, 150–170. doi: 10.1080/17408989.2016.1157574
- Roose, I., Goossens, M., Vanderlinde, R., Vantighem, W., and Van Avermet, P. (2018). Measuring professional vision of inclusive classrooms in secondary education through video-based comparative judgement: an expert study. *Stud. Educ. Eval.* 56, 71–84. doi: 10.1016/j.stueduc.2017.11.007
- Ryan, S., and van der Mars, H. (2022). Can you hear me now ...? *J. Phys. Educ. Recreat. Dance* 93, 22–26. doi: 10.1080/07303084.2021.2000528
- Schicklinski, K., and Reuker, S. (2023). Der Professionelle Blick von Lehrkräften auf Schüler*innenunterschiede: Ein Scoping Review nationaler und internationaler Forschungsarbeiten. *Z. Sportpäd. Forsch.* 11, 136–157. doi: 10.5771/2196-5218-2023-2-136.
- Schlag, S., and Glock, S. (2019). Entwicklung von Wissen und selbsteingeschätztem Wissen zur Klassenführung während des Praxissemesters im Lehramtsstudium. *Unterrichtswiss.* 47, 221–241. doi: 10.1007/s42010-019-00037-8
- Schuldheisz, J. M., and van der Mars, H. (2001). Active supervision and students' physical activity in middle school physical education. *J. Teach. Phys. Educ.* 21, 75–90. doi: 10.1123/jtpe.21.1.75
- Seidel, T., Kosel, C., Böheim, R., Gegenfurtner, A., and Stürmer, K. (2024). “A cognitive perspective on teachers' professional vision: how teachers' professional knowledge shapes a professional vision” in *Teacher professional vision: theoretical and methodological advances*. eds. A. Gegenfurtner and R. Stahnke (London: Routledge), 50–66.
- Seidel, T., and Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *Am. Educ. Res. J.* 51, 739–771. doi: 10.3102/0002831214531321
- Sherin, M. G., Richards, J., and Larison, J. (2024). “Professional vision and teacher noticing looking back and looking forward” in *Teacher professional vision: Theoretical and methodological advances*. eds. A. Gegenfurtner and R. Stahnke (London: Routledge), 67–83.
- Sherin, M. G., and van Es, E. A. (2008). Effects of video club participation on teachers' professional vision. *J. Teach. Educ.* 60, 20–37. doi: 10.1177/0022487108328155
- Stahnke, R., and Blömeke, S. (2021). Novice and expert teachers' situation-specific skills regarding classroom management: what do they perceive, interpret and suggest? *Teach. Teach. Educ.* 98:103243. doi: 10.1016/j.tate.2020.103243
- Stahnke, R., Schueler, S., and Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decisionmaking: a systematic review of empirical mathematics education research. *ZDM* 48, 1–27. doi: 10.1007/s11858-016-0775-y
- Stevenson, N. A., VanLone, J., and Barber, B. R. (2020). A commentary on the misalignment of teacher education and the need for classroom behavior management skills. *Educ. Treat. Child.* 43, 393–404. doi: 10.1007/s43494-020-00031-1
- Stokking, K., Leenders, F., De Jong, J., and Van Tartwijk, J. (2003). From student to teacher: reducing practice shock and early dropout in the teaching profession. *Eur. J. Teach. Educ.* 26, 329–350. doi: 10.1080/0261976032000128175
- van der Mars, H., Darst, P., Vogler, B., and Cusimano, B. (1994). Active supervision patterns of physical education teachers and their relationship with student behaviors. *J. Teach. Phys. Educ.* 14, 99–112. doi: 10.1123/jtpe.14.1.99
- van Driel, S., Wolff, C. E., Crasbon, F., Brand-Gruwel, S., and Jarodzka, H. (2022). A coding scheme to clarify teachers' interactive cognitions in noticed classroom management situations from an actor's perspective. *Teach. Teach. Educ.* 111:103602. doi: 10.1016/j.tate.2021.103602
- van Es, E. A., and Sherin, M. G. (2002). Learning to notice: scaffolding new teachers' interpretations of classroom interactions. *J. Technol. Teach. Educ.* 10, 571–596.

- Weber, K. E., Gold, B., Prilop, C. N., and Kleinknecht, M. (2018). Promoting pre-service teachers' professional vision of classroom management during practical school training: effects of a structured online- and video-based self-reflection and feedback intervention. *Teach. Teach. Educ.* 76, 39–49. doi: 10.1016/j.tate.2018.08.008
- Weber, K. E., Neuber, K., and Prilop, C. N. (2023). Videobasierte Reflexion von klassenführungsspezifischen Ereignissen – Welche Rolle spielen Wissen und Reflexionsbereitschaft von Lehramtsstudierenden? *Z. Erziehungswiss.* 26, 1235–1257. doi: 10.1007/s11618-023-01195-3
- Weber, K. E., Prilop, C. N., Viehoff, S., Gold, B., and Kleinknecht, M. (2020). Fördert eine videobasierte Intervention im Praktikum die professionelle Wahrnehmung von Klassenführung – Eine quantitativ-inhaltsanalytische Messung von Subprozessen professioneller Wahrnehmung. *Z. Erziehungswiss.* 23, 343–365. doi: 10.1007/s11618-020-00939-9
- Weyers, J., König, J., Santagata, R., Scheiner, T., and Kaiser, G. (2023). Measuring teacher noticing: a scoping review of standardized instruments. *Teach. Teach. Educ.* 122:103970. doi: 10.1016/j.tate.2022.103970
- Weyers, J., Kramer, C., Kaspar, K., and König, J. (2024). Measuring pre-service teachers' decision-making in classroom management: a video-based assessment approach. *Teach. Teach. Educ.* 138:104426. doi: 10.1016/j.tate.2023.104426
- Witt, J., Schorer, J., Loffing, F., and Roden, I. (2024). Eye-tracking research on teachers' professional vision: a scoping review. *Teach. Teach. Educ.* 144:104568. doi: 10.1016/j.tate.2024.104568
- Wolff, C. E., Jarodzka, H., and Boshuizen, H. P. A. (2017). See and tell: differences between expert and novice teachers' interpretations of problematic classroom management events. *Teach. Teach. Educ.* 66, 295–308. doi: 10.1016/j.tate.2017.04.015
- Wolff, C. E., Jarodzka, H., and Boshuizen, H. P. A. (2021). Classroom management scripts: a theoretical model contrasting expert and novice teachers' knowledge and awareness of classroom events. *Educ. Psychol. Rev.* 33, 131–148. doi: 10.1007/s10648-020-09542-0
- Wolff, C. E., Jarodzka, H., van den Bogert, N., and Boshuizen, H. P. A. (2016). Teacher vision: expert and novice teachers' perception of problematic classroom management scenes. *Instr. Sci.* 44, 243–265. doi: 10.1007/s11251-016-9367-z