

Baumgartner, Matthias; Jeisy, Eric; Berthold, Clemens

From knowledge to performance in physical teacher education: A Delphi study and a pretest for the content validation of the test instruments

Schweizerische Zeitschrift für Bildungswissenschaften 45 (2023) 2, S. 151-163



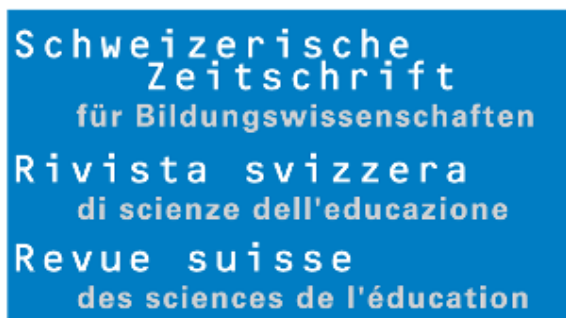
Quellenangabe/ Reference:

Baumgartner, Matthias; Jeisy, Eric; Berthold, Clemens: From knowledge to performance in physical teacher education: A Delphi study and a pretest for the content validation of the test instruments - In: Schweizerische Zeitschrift für Bildungswissenschaften 45 (2023) 2, S. 151-163 - URN: urn:nbn:de:0111-pedocs-277930 - DOI: 10.25656/01:27793; 10.24452/sjer.45.2.6

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-277930>

<https://doi.org/10.25656/01:27793>

in Kooperation mit / in cooperation with:



<http://www.rsse.ch/index.html>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

From knowledge to performance in physical teacher education: A Delphi study and a pretest for the content validation of the test instruments

Matthias Baumgartner, PH St. Gallen
Eric Jeisy, PH St. Gallen
Clemens Berthold, PH St. Gallen

The basis for evaluating the effectiveness of interventions in the education and continuing professional development of physical education (PE) teachers involves valid, reliable, and objective instruments, which are largely lacking. Content validity is a necessary basis for the development of such instruments. Accordingly, this study presents the content validation of a test for the assessment of (classroom management-related) professional knowledge and the perception, interpretation, and decision-making (PID) of prospective PE teachers. The result is a content-validated knowledge test (110 test items) as well as a video-based test for the assessment of PID (ten vignettes, 217 items). In a next step, the instruments are empirically validated.

1. Test instruments as a starting point for examining the effectiveness of the education and continuing professional development of physical education teachers

The consolidation of data showing the high (potential) influence of physical education (PE) teachers on the development of students has led to a growth in studies on the effectiveness of teacher education and continuing professional development of (PE) teachers. In the past two decades, a competence-oriented shift has been observed (e.g., Blömeke et al., 2015), which has also found its way into PE (teacher) research (e.g., Baumgartner, 2018a; Meier, 2020). When it comes to evaluating the effectiveness of various higher education teaching and learning arrangements in the education and continuing professional development of PE teachers, the data entry within the framework of the competence-oriented (PE) research tradition is limited (e.g., Baumgartner, 2022b). Valid, reliable, and objective test instruments are the basis for examining the effectiveness of the education and continuing professional development of PE teachers, but these are largely lacking in PE teacher research (e.g., Wibowo & Heemsoth, 2019). Even if the focus is on the classroom management (CM) of PE teachers, which is undisputedly a relevant factor in good PE teaching (Baumgartner et al., 2020), there are no test instruments available for stringently capturing the three CM-related competence facets (professional competency; situated perception, interpretation and decision-making abilities (PID); performance), cf. Section 2.1; cf. (Baumgartner, 2022a). The criterion of content validity should be regarded as a necessary basis for the development of such test instruments. When the content of a test can acquire a construct exhaustively (and faithfully) with regards to the most important dimensions (latent variable), content validity is confirmed as high (Döring & Bortz, 2016). The article addresses this gap. Specifically, it focusses on the content validation of a CM-related knowledge test, as well as a video vignette-based test to determine the quality of situated and CM-related perception, interpretation, and decision-making processes realized within the framework of the project “From Knowledge to Performance in Physical Education: Prospective PE-teachers’ transformation of competences (WiPe-Sport)”¹, supported by the Swiss National Science Foundation (SNSF). The aim is to create a (necessary) foundation for evaluating the effectiveness of the education and continuing professional development of PE teachers.

¹ For further information, see <https://p3.snf.ch/project-192397>

2. Theory

2.1 Project Framework: On the presumption of correlation and development in the integrative understanding of competence – a desideratum

Within the framework of the competence-oriented research paradigm, an integrative understanding of professional competence has been established (Baumgartner, 2022a). It is assumed that “competence is viewed along a continuum from traits that underlie perception, interpretation, and decision-making skills, which in turn give rise to observed behavior in real-world situations” (Blömeke et al., 2015, p. 3). In this integrative understanding, the term professional competence “is used to describe the whole construct of competence within a profession” (Baumgartner, 2022a). The construct of professional competence can be divided into three competence facets: 1) the facet of professional competency (dispositions) as a latent construct consisting of different aspects of competency (e.g., professional knowledge, motivational orientations, beliefs and values; Baumert & Kunter, 2006; Blömeke et al., 2015; Baumgartner, 2018a); 2) the facet of situation-specific abilities with the three sub-facets of perception, interpretation and decision-making (PID); 3) the facet of performance (in the sense of skills; cf. Blömeke et al., 2015; Baumgartner, 2022a; Shavelson, 2013). In this regard, it is assumed that the correlation between professional competency (dispositions) and performance is mediated by the situated PID abilities. It is suspected that the three competence facets are (positively) correlated (Blömeke et al., 2022) and that PE teachers who have a higher quality of professional competency or the PID perform better in a real-world situation than others (Baumgartner, 2017). Furthermore, the developmental presumption is that improving competence area-related performance (e.g., implementing good CM in PE) can be accomplished by a) enhancing professional competency, respectively the aspects of competency (e.g., advancing the quality of CM-related knowledge through literature-based work), b) improving PID (e.g., increasing the quality of CM-related PID through reflecting on CM-related videos), and c) through deliberative practice of the application of the quality dimensions (e.g., the implementation of the CM-related dimension of monitoring; Baumgartner, 2022a). Although these presumptions of correlation and development also form the basis of the PE teacher education (PETE) and continuing professional development of PE teachers, the data in the field of professional studies can be described as inconsistent (Blömeke et al., 2022).

2.2 Facet of performance: Classroom management in PE

CM can be described as the sum of teachers’ actions that aims to create a conducive learning environment for cognitive, social-emotional, and motor learning (Baumgartner et al., 2020). The quality of PE teachers’ CM can ultimately be measured on a performative level (facet of performance) by examining what actions PE teachers take in the complex situation of PE (Baumgartner et al., 2020). Various instruments are available for measuring the quality of CM-related performances (e.g., Baumgartner et al., 2020; Heemsoth, 2014). Baumgartner et al. (2020) present an observation instrument (KlaPe-Sport), which, in comparison to others, gives a good representation of a PETE and continuing professional development-related understanding of the quality of CM. CM is accordingly recorded broadly and is thus intended to meet the quality understanding of requirement-related fidelity (Baumgartner, 2022b). The instrument consists of nine dimensions (latent variables) and 27 items (manifest variables) (cf. Tab. 1) and represents the quality understanding of good CM in this paper.

Table 1

An instrument for measuring classroom management-related performances in PE teaching (Baumgartner et al., 2020)

Dimensions	Items (n)	Example of item. The PE teacher...
Monitoring	4	...chooses positions in the room from which she/he has a good overview of what is going on in the class.
Safety	3	...organizes the learning environment safely (e.g., use of gym mats).
Clarity of announcements	3	...formulates objectives/tasks/exercises in a comprehensible way.
Dealing with disruptions	3	...manages to focus the learners’ attention on the lesson in the case of minor disturbances.
Use of material	3	...does not use too much equipment and materials in PE lessons.
Momentum	3	...progresses quickly in terms of content so that there is no waste of time.
Smooth transitions	3	...intervenes in a targeted way when transitions do not run smoothly.
Overlapping	3	...can dedicate himself/herself to a teaching event without neglecting the class.
Group mobilization	2	...enables many learners to learn at the same time.

To be able to investigate the presumption of correlations and development (cf. Section 2.1) considering the understanding of the quality of good CM performance as the dependent variable (DV) (cf. Tab. 1), test instruments are needed that can stringently record a) the quality of CM-related knowledge (independent variable, IV) – interpreted here as the primary aspect of competency (Baumert & Kunter, 2006) – and b) the CM-related PID abilities of PE teachers (mediator variable, MV).

2.3 Facet of professional competency: knowledge of CM as an aspect of competency

The facet of professional competency represents a latent construct consisting of different aspects of competency (e.g., professional knowledge, motivational orientations, beliefs and values; Baumgartner, 2022a). Professional competency is defined as “the knowledge, cognitive skills, and affective-motivational dispositions” (Blömeke et al., 2015, p. 4) which are responsible for selecting, maintaining, and supervising professional actions (Baumert & Kunter, 2006). Professional knowledge – interpreted here as the primary aspect of competency (Baumert & Kunter, 2006) – is regarded as a necessary but insufficient prerequisite for the other facets (Baumgartner, 2017).

Conceptually, professional knowledge in (PE) teacher research is usually categorized using the three dimensions of content, pedagogical content, and general pedagogical knowledge (Baumgartner, 2017). Knowledge about CM can be categorized within pedagogical knowledge (Voss et al., 2015). In addition to the typology of professional knowledge, various forms of the representation of knowledge are also distinguished, which play different roles in teaching operations. (Explicit) declarative knowledge refers to articulable knowledge about facts or events, which can be acquired through theory-based research or in practice (Guerriero, 2017). The (mostly implicit) procedural knowledge enables the routine execution of action sequences (procedures) to achieve a specific goal (Guerriero, 2017) and can therefore only be derived from performance (Neuweg, 2014).

Regarding the state of research on pedagogical knowledge, findings show that it has a moderate positive effect on teaching quality and student outcomes (Ulferts, 2019). Various validated instruments are available for assessing pedagogical knowledge (Voss et al., 2015), which differ in their content focus and terminology, as well as methods, in addition to the form of the knowledge and contextualization (Brühwiler & Hollenstein, 2021). Situated instruments in particular provide additional information content that goes beyond the effects of pure knowledge (Kaiser et al., 2017). The situational abilities contribute to the measured effects to the same extent as pedagogical knowledge, while presumably also playing an important role in its application (Ulferts, 2019). Accordingly, it is deduced that knowledge is captured in an uncontextualized manner (e.g., CM-related declarative knowledge; Baumgartner, 2018b). In terms of the declarative components of pedagogical knowledge, a small, statistically significant effect on classroom performance was identified, but it was found to be completely mediated by procedural knowledge (Lenske et al., 2016). Nonetheless, declarative knowledge appears to be important in the formation of procedural knowledge (Jeschke et al., 2021).

If the focus is directed to CM-related knowledge, the available uncontextualized instruments on pedagogical knowledge tend to depict the construct of CM only superficially (Römer & Rothland, 2015). Specifically, there are no instruments available that are oriented (in the sense of an integrative understanding of competence) to an observation instrument for measuring performance. In this sense, the present work intends to address this gap.

2.4 Facet of professional competency: situation-specific ability of perception, interpretation, and decision-making (PID)

The ability of teachers to recognize or focus on important events in the classroom that are relevant to learning and teaching, to interpret these based on knowledge and experience and, if necessary, to derive decisions for action from them, is considered a central facet in the integrative understanding of competence (Blömeke et al., 2015; Gold et al., 2021). Different concepts and terms are used for these situation-specific abilities between professional competency and performance. “professional Vision” (e.g., Reuker, 2017a, 2017b), “(teacher) noticing” (e.g., Langer, 2022) or “situation-specific skills” (e.g., Blömeke et al., 2015) are only three examples of numerous conceptions, which often refer to the same authors as a basis (e.g., van Es & Sherin, 2002), but sometimes adopt different theoretical perspectives (König et al., 2022). Essentially, a distinction can be made between holistic and analytical conceptions. Holistic approaches consider and measure situational (cognitive) abilities as an overall construct (e.g., Roose et al., 2018). In analytical approaches, two (e.g., van Es & Sherin, 2002), often three (e.g., Blömeke et al., 2016) and sometimes four (e.g., Junker et al., 2021) cognitive sub-processes are distinguished and recorded separately in a more empirical way (Gippert et al., 2022).

Within the conceptional framework of the project (Baumgartner, 2022a), we are guided by the work of Blömeke et al. (2015, 2016, 2022) in operationalizing situation-specific abilities. They distinguish between

the three sub-facets of the *perception* of situations relevant to learning and teaching, the analysis and *interpretation*, and the subsequent *decision-making* about options for action. In terms of CM in the subject area of PE, this means, for example, that PE teachers must be able to recognize the safety-relevant classroom behavior of students, interpret this in a knowledge-based manner, and be able to decide in situ on a suitable CM-related behavior.

In video-based assessments, scripted or real video vignettes are used as stimuli to be able to record situational abilities with questions or tasks related to them. Compared to paper-and-pencil tests, they offer the advantage that they better reflect the complexity of the classroom situation (König et al., 2021). Test instruments with close-ended questions usually use rating items in which the subjects' answers are compared with the results of an expert rating to quantify test performance (for CM, e.g., Gold et al., 2017; Seidel & Stürmer, 2014).

If the focus is directed to CM-related PID in PE, no specific instruments are available. This research gap is to be addressed in the present paper.

3. Methods

In the following section, the conceptual test developments for assessing the CM-related knowledge and the PID of prospective PE teachers are described.

3.1 Test development

In a first step, the first version of the test is the result of systematic literature analysis, an intensive professional discussion within the project team, as well as with the involvement of external SNSF project partners (experts). The content framework is provided by the nine dimensions of the CM-related observation instrument validated by Baumgartner et al. (2020; cf. Tab. 1). This is the common reference for all test items. A closed-response format with dichotomous single-choice items was chosen in order to utilise the specific advantages of the item response theory (IRT) for the (later) empirical construct validation (Clark & Watson, 2019).

3.1.1 Test to assess CM-related knowledge

CM-related knowledge is to be captured in an uncontextualized manner within the framework of the SNSF project and in consideration of the integrative competence structure model (Baumgartner et al., 2020; Blömeke et al., 2015), so that the selectivity for the situated PID is guaranteed (cf. Section 2.4). The professional knowledge base about CM was narrowed down to the declarative type, to meet the model-based assumptions (cf. Section 2.1) and practice-related content, and increase reliability between the instruments. It is compiled from the best available empirical findings (Knogler et al., 2022) on effective CM, as well as normative criteria of good CM (Berliner, 2001). Meta-analyses (e.g., Hattie, 2009), (systematic) reviews (e.g., Evertson & Weinstein, 2006), original research papers (e.g., van der Mars et al., 1994), criteria for good CM (e.g., Ophardt & Thiel, 2013), or recommendations from practice-oriented literature (e.g., Söll & Kern, 2008) served as the basis for the development of dichotomous items (e.g., "To monitor the classroom, a teacher should [...]"), for which several dimension-specific single-choice items were formulated (e.g., "[...] choose a fixed position that allows him/her to keep all students in sight."). Each item describes a single purpose of action, which is to be evaluated by the test-takers as correct or incorrect.

This literature review was used to identify evidence of existing practices for each of the nine dimensions and detect conceptual differences within them (e.g., monitoring: active supervision, withitness, oversight). These subdimensions were deduced in an effort to further structure the test by offering different item prompts within each category. Starting with one subdimension per item of the observation instrument, some categories were later combined based on theoretical assumptions. In the case of monitoring, "active supervision" (van der Mars et al., 1994) describes the teachers' "physical movement around the teaching space" (Cothran & Kulinna, 2015, p. 242) (e.g., location, rate of movement), "withitness" (Kounin, 1970), refers to the teacher-student interaction as teachers "demonstrated awareness through the timeliness and accuracy of desists" (Doyle, 2006, p. 109), and "situational awareness/overview/monitoring" depicts the way teachers scan the room to attend to a single student's actions while monitoring the group as a whole.

In the design of the knowledge test, definitions of the specific terms used were included where they occurred, and presented to the participants alongside the items. This was to prevent errors resulting from unfamiliarity with our specific definitions of the dimensions or certain concepts (e.g., "small disruptions", cf. Tab. 1).

3.1.2 Video-based test to assess situation-specific CM-PID abilities

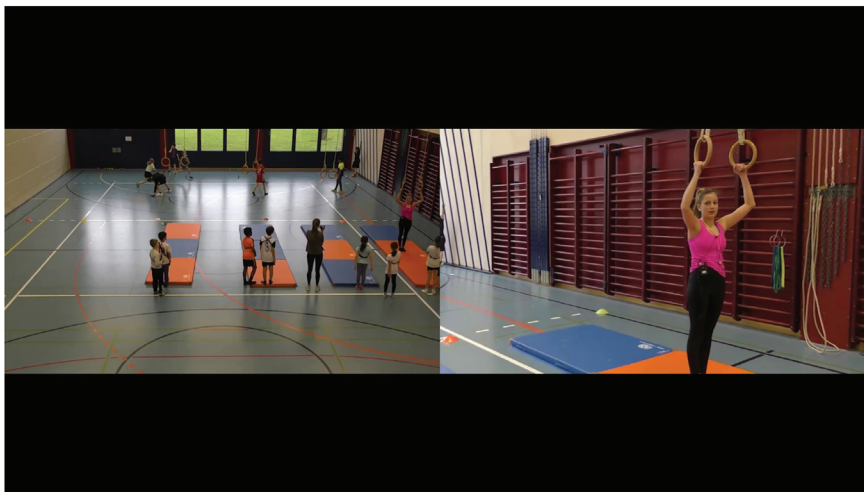
As explained in Section 2.4, video-based assessments have proven to be the gold standard for assessing situation-specific abilities, as they are better able to reflect the complexity and authenticity of challenging teaching situations than other assessments (König et al., 2022). We also follow this practice and were able to draw on a data pool of 60 videotaped PE lessons of pre-service teachers at the kindergarten and elementary school level to develop the video-based test. The videotaped lessons were each filmed with two guided cameras. A wireless microphone was used to record the teachers' statements.

To record CM-related sub-facet *perception*, a procedure is used in which the test participants watch a teaching sequence once, and then have to indicate which dimension of CM they perceived as relevant in the video sequence observed. An introductory text points out that this can be particularly good and/or lead to improvable behavior in the area of CM. In addition, short definitions of CM are displayed to the participants for each video vignette to ensure that the concepts guiding the research are being tested (Gippert et al., 2022). Selected basic dimensions of CM PE (Baumgartner et al. 2020) are presented as response options (cf. Fig. 1).

Figure 1

Screenshot of a video vignette representing monitoring and safety as teaching-relevant dimensions of good classroom management²

Please watch the following video once. Afterward you will have to answer different questions on this video vignette.



In this video, which aspect or aspects of classroom management would you consider relevant?

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

For the PID sub-facet of *interpretation*, the answer must be inferred from an observation based on the relevant, situated CM dimension in each case (Hellermann et al., 2015). To meet this requirement, two to three items were formulated for each relevant CM dimension represented in the video. After analyzing and interpreting the teaching sequence, respondents had to judge whether the CM actions described in the item are correct or not: “Which of the following CM-related teachers’ actions happened?” (example item monitoring, see Figure 1: “The teacher concentrates on the group doing gymnastics on the rings without losing sight of the class”).

² In addition to monitoring and safety, the test participants are offered four other CM-related dimensions as distractors for the facet of perception, and must decide whether they perceived each aspect in the video sequence as particularly relevant for CM.

Decision-making is assessed as an evaluation of options for teaching behavior, following Blömeke et al. (2016). For this purpose, the test participants are asked: “If you were the teacher in this situation, what would you do differently to improve CM?”. As answer options, two or three behavior alternatives are presented for each CM dimension represented, and the test participants must decide for each whether they consider a corresponding procedure to be goal-oriented in the specific situation (example item monitoring, see Figure 1: “When instructing and demonstrating, I position myself in such a way that I also keep an eye on the small group playing”). As in the other two cognitive sub-facets of PID, a master rating serves as a reference standard for assessing whether or not the action option is goal-oriented, which will be based on expert consensus. When developing the items for the three cognitive processes of the PID, care was taken to ensure that the items for the various aspects of CM were evenly distributed across all the dimensions (cf. Tab. 3), and that the item pool was large enough to enable a response in particular to the quantitative adaptations within the framework of the content validation through the Delphi study and the pretest (Wibowo & Heemsoth, 2019).

3.2 Delphi studies

To validate the content of the tests, two separate three-phase Delphi studies (Häder, 2014) were organized, one for the knowledge test and one for the PID test, and conducted in the same way. A Delphi study to develop solutions to complex problems from the individual contributions of the experts involved (Häder, 2014). For the present study, the ideal participants are experts who have teaching experience in the area of PE, as well as having been active as lecturers and/or researchers in the field of PETE. This should address a group of experts who have a professional understanding of good classroom management in physical education.

The experts who participated in the two Delphi studies have teaching experience in the field of physical education (knowledge test: $M = 12.1$ years, $SD = 6.2$, PID: $M = 16$ years, $SD = 11.3$) and have worked as lecturers and/or researchers in the field of PETE ($M = 10.7$ years, $SD = 4.5$, $M = 14.5$, $SD = 11.4$). Seven experts (3 = female (f.), 4 = male (m.), 0 = diverse (d.)) participated in the Delphi study for the content validation of the knowledge test and eight (3 = f., 5 = m., 0 = d.) participated in the validation of the PID test.

In an initial Delphi round, the experts completed either the knowledge or PID test. In addition, they were asked to provide written comments.

In the second Delphi round, the participants were presented with individual and aggregated results, together with the correct and incorrect answers determined by the project team. The outcomes were then discussed by item in small groups, each moderated by a project member. The focus was on items where there was little agreement among the experts. This step served to check professional relevance, correctness, content representativeness, linguistic comprehensibility, and unambiguity, as well as the categorization of the test instruments from the experts' point of view (Moosbrugger & Kelava, 2020). The discussion results of the second Delphi round led to the project team revising and reducing the items. In a third Delphi round, the revised instruments were presented to the experts again for renewed expert review and consensus building (Häder, 2014), and were approved.

3.3 Pretest with target group

In the next step, the test instruments validated by the experts were transferred to a pretest survey with 54 participants. The aim was to assess the clarity and comprehensibility of the instruments from the test participants' point of view (Moosbrugger & Kelava, 2020) and to generate an initial overview of the difficulty of the items. Kindergarten and primary education students from two Swiss universities of teacher education participated in the survey (response rate 62%, 47 = f., 7 = m., 0 = d.; age $M = 24.24$, $SD = 6.21$). The mean completion time of the total test was 44 minutes ($SD = 28$; knowledge test: $M = 8$ min, $SD = 6$ min; PID test: $M = 28$ min, $SD = 19$ min). The pretest was conducted as an online survey using the survey tool Limesurvey (version 5.2.1). The knowledge and the video-based PID tests were supplemented by a general question section regarding the main survey. Open questions on the clarity of the content and the technical procedure of the test, as well as feedback options on the overall survey, complemented the online survey. To reduce the workload for the test participants, both the knowledge and the PID tests were halved and participants were randomly assigned to one half of each test. Based on the objective of the pretest, the results were evaluated quantitatively based on the probability of solution (relevance, clarity), and qualitatively based on the students' feedback (clarity).

4. Results

4.1 CM-knowledge test

The first version of the test (the result of systematic literature analysis, an intensive professional discussion within the project team, as well as with the involvement of external SNSF-project partners) consists of 20 subdimensions with a total of 175 items classified by the nine dimensions of good CM (cf. Tab. 1).

In the Delphi study, 59 items were revised, 5 items were newly developed, and 63 items were removed.

The pretest with the target group resulted in a high solution rate of the items ($M = 78\%$, $SD = 23\%$) for the participating students. It is striking that many items (35%; $n = 46$) could be answered correctly by a large proportion of participants ($n > 90\%$). If the probability of solving an item was small ($n < 20\%$), combined with a lower level of agreement among the experts in the Delphi studies, it was concluded that clarity was too low. Thus, seven items were excluded from the knowledge test. Based on the evaluation of the open questions, unclear terms were replaced or explained (e.g., momentum). Definitions were supplemented or revised (e.g., small disturbances). However, no more changes were made to the content of the remaining items, or to the number of dimensions or subdimensions. There were no significant technical problems during the test.

As an outcome, the tests for measuring CM-related knowledge resulting from the described content validation steps are presented and described below (cf. Tab. 2).

Table 2

Numbers and distribution of items in the structure of classroom management-related knowledge test

Dimension	Sub - dimension	Items	Solution rate % <i>M (SD)</i>	Example of item
monitoring	3	13	85 (10)	To monitor the classroom, a teacher should... ...choose a fixed position that allows him/her to keep all students in sight. (false)
dealing with disruptions	3	14	62 (33)	In order to protect the primary teaching vector, off-task behavior... ...should be stopped before other students catch on. (true)
clarity of instructions	3	14	75 (20)	For clarity of instructions a teacher should... ... routinely repeat all rules or tasks of activities, even if students should be familiar with them. (false)
group mobilization	2	12	87 (12)	To facilitate learning for all students simultaneously, teachers should... ...provide the same tasks to all students. (false)
momentum	2	12	69 (20)	To enable high momentum a teacher should... ...progress briskly to prevent intermissions. (true)
overlapping	1	6	68 (35)	In case of overlapping class events, a teacher should... ... handle them simultaneously. (true)
smooth transitions	2	10	81 (25)	When transitioning between different classroom activities, a teacher should... ...stop the current activity, before taking further actions. (true)
safety	2	18	77 (21)	To run safe PE lessons, a teacher should... ...not allow students to enter the gym's stockroom without supervision. (false)
use of material	2	11	93 (10)	When using sports equipment, a teacher should... ...prevent use for which it is not specified. (false)

The CM knowledge test resulting from the content validation, which is oriented towards the nine dimensions of the CM observation instrument (cf. Tab. 1), consists of 110 items that are collected using 20 subdimensions. The distribution (number of items per dimension) is not uniform. The unequal distribution is related to the different information content of the diverse (and observational instrument-related) dimensions. Accordingly, it was to be expected that the monitoring dimension, for example, would have a larger item pool compared to the smooth transitions dimension. Furthermore, there is no empirical evidence to show the scale of the influence of the different dimensions of good classroom management on the quality of physical education. Consequently, we accept an unequal distribution of items with respect to the different dimensions. Changes during the Delphi study led mainly to linguistic adjustments, while the content conception largely withstood the critical eye of the

experts. The relatively large differences in the solution rate during the pretest between individual dimensions can be put into perspective with closer scrutiny. Dimensions with lower solution rates contained the terms that were described as unclear, and therefore revised. Whether the outcomes of the pre-pilot will be substantiated will only be shown in the pilot for empirical validation.

4.2 PID test

From the data pool of 60 videotaped PE lessons of pre-service teachers, a total of ten video vignettes with a duration of 1.19 to 3.27 minutes were selected. To be able to depict the construct of CM regarding the most important dimensions in a valid, i.e., exhaustive, manner (Wibowo & Heemsoth, 2019), care was taken to ensure that all nine dimensions regarding CM-related performance (cf. Tab. 1) were represented at least twice in different CM-related situations (video vignettes).

In the context of the Delphi study, 24 items were revised, three items were newly formulated, and 24 items were deleted.

The items of the PID test also showed a high solution rate ($M = 73\%$, $SD = 20\%$) among the students who participated in our pretest (Clark & Watson, 2019). A high percentage of items in the PID test (23%; $n = 50$) could be answered correctly by a large proportion of participants ($n > 90\%$). Again, items with a low solution rate ($n < 20\%$) were checked for lower level of agreement among the experts in the Delphi study. Clarity was deemed too low if that coincided. Subsequently, five items were excluded from the PID test. Based on the evaluation of the open questions, unclear terms were replaced or explained. The remaining items' content and the quantity of video vignettes, however, remained unchanged. During the test, there were no serious technical issues. As an outcome, the tests for measuring CM-related PID resulting from the content validation steps described are presented below (cf. Tab. 3).

Table 3

Video-based test for measuring classroom management-related situation specific PID abilities in PE teaching

Vignette	Duration (min:s)	Subject-specific content	Dimensions ³	Items (n)	Solution rate % <i>M (SD)</i>
1	01:19	Gymnastics	1,2	22	77 (23)
2	01:34	Dance	3,7	21	78 (17)
3	01:35	Gymnastics	9,5	21	77 (21)
4	02:15	Games	4,1	18	73 (17)
5	01:33	Circus	8,6,1	22	73 (17)
6	02:01	Games	3,7,8	22	73 (21)
7	03:27	Games	1,6	19	77 (22)
8	01:31	Gymnastics	4,9	22	75 (20)
9	01:44	Gymnastics	5,2,3	23	69 (21)
10	02:54	Games	9,8,6,1,7	27	68 (22)

After content validation, the PID test consists of ten video vignettes that depict different subject content. Each dimension is represented at least twice in a teaching sequence and the three situation-specific sub-facets of the PID are recorded by means of 217 items. The number of items is evenly distributed across the cognitive sub-facets (perception = 70 items; interpretation = 72 items; decision-making = 75 items) and the different CM-dimensions (ranging from 16 to 20 items per dimension). The average solution rates per vignette can generally be described as high (Clark & Watson, 2019), whereby the instructional sequence with the most complex content (vignette ten with five dimensions) also shows the lowest solution rate. Overall, the video vignettes differ only slightly from each other in terms of item difficulty.

³ Monitoring (1); Safety (2); Clarity of announcements (3); Dealing with disruptions (4); Use of material (5); Momentum (6); Smooth transitions (7); Overlapping (8); Group mobilization (9).

5. Discussion and conclusion

Valid, reliable and objective test instruments, which are largely lacking in the context of PE (teacher) research, form the basis for evaluating the effectiveness of teaching and learning arrangements in the PETE and continuing professional development of PE teachers. In this respect, the present work should contribute to addressing the research gap outlined. Specifically, considering an integrative understanding of professional competence (Baumgartner, 2022a) and following a validated observation instrument for measuring CM-related performance (facet of performance), a test for recording CM-related knowledge (facet of professional competency) and a video vignette-based test for measuring CM-related PID (facet of PID) were validated in terms of content. The content validation of test instruments is a prerequisite for the recording of variables. In addition, a way of transferring the same CM-related content into different tests is demonstrated. The test-related items were methodically derived from the literature, condensed by the project team and discussed with experts. In three-phase Delphi studies, the test content of the two prototypes were validated by lecturers and/or scholars in the field of PETE. Subsequently, the content of the existing tests was validated in the sense of pre-pilot testing.

With the development of the knowledge test and the PID test, two comprehensive instruments are available, the dimensions and items of which - from the point of view of the experts (Delphi study) - are relevant, correct and representative in term of content and linguistic comprehensive. The pretest also showed that both tests are clear and comprehensible from the respondents' point of view and that there were no significant problems when administering the test.

As limitations, it should be mentioned that regarding the knowledge test, the narrow choice in the form of declarative knowledge about effective practices has been justified within the model-based assumptions, but is discussed critically as a technological approach to pedagogics and a knowledge base that seems to evade completion (Herzog, 2016).

In addition, the two tests (pretest) were solved by pre-service teachers, therefore it is not clear how difficult the test is for in-service teachers.

Furthermore, the test was developed for the kindergarten and primary school. Consequently, the tests cannot be used directly to measure the quality of classroom management-related knowledge or the PID at other levels of education.

The close-ended recording of PID is not uncontroversial and, especially with regard to the two sub-facets perception and decision-making, the depth of content that could be achieved with an open-ended question format cannot be achieved (Gippert et al., 2022). It is also still open as to whether the conceptual separation of the situation-specific abilities into three different sub-facets can be empirically confirmed in the structural test validation. Although this has been achieved several times in work in empirical educational research (e.g., Blömeke et al., 2016), it is still unknown whether this may succeed in combining a domain-specific approach (CM) with a PE-specific perspective.

From a test economy perspective, the size of the test instruments is also critical and the high solution rates, which are problematic regarding the main study, will have to be reduced in further test development (Clark & Watson, 2019). Possible solutions to these problems include reducing the number of low-difficulty items in the item pool and/or reducing the number of video vignettes in the PID test. Due to the small number of cases in the pretest, we have so far refrained from a large-scale deletion of low-difficulty items. Although that would lead to higher test difficulty, it could challenge the content validity.

In order to empirically validate the content presented here by means of construct and criterion validation (Döring & Bortz, 2016), a pilot study with 995 participants is currently being conducted and evaluated. We hope to be able to present valid test instruments after the pilot study, which will make it possible for the first time to test the integrative competence structure model in the subject area of PE in relation to training across three competence facets. However, in terms of examining the entire chain of effects up to the quality of teaching and the performance of the students (Blömeke et al., 2022), there remains a lack of empirically validated and generally accepted test instruments for the user-side recording of learning performance in the subject area of PE. There is still a great need for research on this topic.

References

- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Baumgartner, M. (2022a). Professional competence(s) of physical education teachers: terms, traditions, modelling and perspectives. *German Journal of Exercise and Sport Research*, 52(4), 550–557. <https://doi.org/10.1007/s12662-022-00840-z>
- Baumgartner, M. (2022b). Professionelle Kompetenz(en) von Sportlehrkräften – Begriffe, Traditionen, Modellierungen und Perspektiven. In R. Sygusch, J. Hapke, S. Liebl & C. Töpfer (Eds.), *Kompetenzorientierung im Sport. Grundlagen, Modellentwurf und Anwendungsbeispiele* (pp. 27–41). Hofmann.
- Baumgartner, M. (2018a). „...Kompetenz ohne Performanz ist leer! Performanz ohne Kompetenz blind...!“ Zu einem integrativen Kompetenzstrukturmodell von Sportlehrkräften. *Zeitschrift für sportpädagogische Forschung*, 6(1), 49–68. <https://doi.org/10.5771/2196-5218-2018-1>
- Baumgartner, M. (2018b). Performanzfortschritt in der Lehrerbildung durch die Arbeit an eigenen video- und textbasierten Unterrichtsfällen? Eine Interventionsstudie zur Verbesserung des Feedbacks bei angehenden Sportlehrkräften. *Zeitschrift für Erziehungswissenschaft*, 21(6), 1135–1155. <https://doi.org/10.1007/s11618-018-0829-8>
- Baumgartner, M. (2017). *Performanzentwicklung in der Ausbildung von Lehrkräften. Eine Interventionsstudie zur Verbesserung des Feedbacks bei angehenden Sportlehrkräften*. Waxmann.
- Baumgartner, M., Oesterheld, V., & Reuter, S. (2020). Konstruktion und Validierung eines multidimensionalen Beobachtungsinstruments zur Erfassung der klassenführungsbezogenen Performanzen von sportunterrichtenden Lehrkräften (KlaPe-Sport). *German Journal of Exercise and Sport Research*, 50(4), 511–522. <https://doi.org/10.1007/s12662-020-00675-6>
- Berliner, D. C. (2001). *Learning about and learning from expert teachers*. *International Journal of Educational Research*, 35(5), 463–482. [https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, 56, 35–46. <https://doi.org/10.1016/j.tate.2016.02.003>
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, 223, 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2022.101600>
- Brühwiler, C., & Hollenstein, L. (2021). The contextualised measuring of general pedagogical knowledge and skills: Exploring the use of knowledge in practice. In H. Ulferts (Ed.), *Teaching as a Knowledge Profession: Studying Pedagogical Knowledge across Education Systems, Educational Research and Innovation*. OECD Publishing. <https://doi.org/10.1787/e823ef6e-en>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427. <https://doi.org/10.1037/pas0000626>
- Cothran, D. J., & Kulinna, P. H. (2015). Classroom management in physical education. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook of classroom management* (pp. 239–260). Routledge.
- Doyle, W. (2006). Ecological Approaches to Classroom Management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97–125). Routledge
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- Evertson, C. M., & Weinstein, C. S. (Eds.). (2006). *Handbook of classroom management: Research, practice, and contemporary issues*. Lawrence Erlbaum Associates Publishers.
- Fenstermacher, G. D. (1994). The Knower and the Known: The Nature of Knowledge in Research on Teaching. *Review of Research in Education*, 20(1), 3–56. <https://doi.org/10.2307/1167381>
- Gippert, C., Hörter, P., Junker, R., & Holodyski, M. (2022). Professional vision of teaching as a focus-specific or focus-integrated skill – Conceptual considerations and video-based assessment. *Teaching and Teacher Education*, 117, 103797. <https://doi.org/10.1016/j.tate.2022.103797>
- Gold, B., Hellermann, C., & Holodyski, M. (2017). Effekte videobasierter Trainings zur Förderung der Selbstwirksamkeit-süberzeugungen über Klassenführung im Grundschulunterricht. *Zeitschrift für Erziehungswissenschaft*, 20(1), 115–136. <https://doi.org/10.1007/s11618-017-0727-5>
- Gold, B., Junker, R., Wissemann, M., Klassen, C., & Holodyski, M. (2021). Are good observers good classroom managers? The relationship between teachers' professional vision and their students' ratings on classroom management. *International Journal of Educational Research*, 109, 101811. <https://doi.org/10.1016/j.ijer.2021.101811>
- Guerriero, S. (Ed.) (2017). *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*. OECD Publishing. <https://doi.org.ezproxy.phsg.ch/10.1787/9789264270695-en>
- Häder, M. (2014). *Delphi-Befragungen. Ein Arbeitsbuch*. Springer.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Heemsoth, T. (2014). Unterrichtsklima als Mediator des Zusammenhangs von Klassenführung und Motivation im Sportunterricht. *Psychologie in Erziehung und Unterricht*, 61(3), 203–215. <https://doi.org/10.2378/peu2014.art16d>
- Hellermann, C., Gold, B., & Holodyski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 97–109. <https://doi.org/10.1026/0049-8637/a000129>
- Herzog, W. (2016). *Kritik der evidenzbasierten Pädagogik*. *Zeitschrift für Erziehungswissenschaft*, 19(1), 201–213. <https://doi.org/10.1007/s11618-016-0711-5>
- Jeschke, C., Lindmeier, A., & Heinze, A. (2021). Vom Wissen zum Handeln: Vermittelt die Kompetenz zur Unterrichtsreflexion zwischen mathematischem Professionswissen und der Kompetenz zum Handeln im Mathematikunterricht? Eine Mediationsanalyse. *Journal für Mathematik-Didaktik*, 42(1), 159–186. <https://doi.org/10.1007/s13138-020-00171-2>
- Junker, R., Gold, B., & Holodyski, M. (2021). Classroom management of pre-service and beginning teachers: From dispositions to performance. *International Journal of Modern Educational Studies*, 5(2), 339–363. <http://dx.doi.org/10.51383/ijonmes.2021.137>

- Jürgens, M. (2021). *Videobasierte Lehrveranstaltungen zur Förderung Professioneller Unterrichtswahrnehmung*. Springer VS.
- Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—cognitive versus situated approaches. *Educational Studies in Mathematics*, 94(2), 161–182. <https://doi.org/10.1007/s10649-016-9713-8>
- Klemenz, S., & König, J. (2019). Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften: Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung. *Zeitschrift für Pädagogik*, 65(3), 355–377. <https://doi.org/10.3262/ZP1903355>
- Knogler, M., Hetmanek, A., & Seidel, T. (2022). Bestimmung und Bereitstellung der „best available“. Evidenz für bestimmte Praxisfelder im Bildungsbereich. In N. McElvany, M. Becker, F. Lauerer, H. Gaspard & A. Ohle-Peters (Eds.), *Optimierung schulischer Bildungsprozesse-What works?* (pp. 135–145) Waxmann.
- König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Nehls, C. F., Musekamp, F., & Kaiser, G. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics*, 107, 189–212. <https://doi.org/10.1007/s10649-020-10021-0>
- König, J., Santagata, R., Scheiner, T., Adleff, A.-K., Yang, X., & Kaiser, G. (2022). Teacher noticing: A systematic literature review of conceptualizations, research designs, and findings on learning to notice. *Educational Research Review*, 36, 100453. <https://doi.org/https://doi.org/10.1016/j.edurev.2022.100453>
- Kounin, J. (1970). *Discipline and group management in classrooms*. Rinehart and Winston.
- Langer, W., Bruns, J., & Erhorn, J. (2022). Entwicklung und Validierung eines videobasierten Testinstruments zur Erfassung des Noticing mit dem Fokus auf Anerkennungsprozesse im inklusiven Sportunterricht. *German Journal of Exercise and Sport Research*, 52, 386–397. <https://doi.org/10.1007/s12662-022-00807-0>
- Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S., & Leutner, D. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft*, 19(1), 211–233. <https://doi.org/10.1007/s11618-015-0659-x>
- Meier, S. (2020). Development and validation of a testing instrument to assess pedagogical content knowledge of German preservice physical education teachers. *Journal of Physical Education and Sport*, 20(5), 3010–3016. <https://doi.org/10.7752/jpes.2020.s5409>
- Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3. ed.). Springer.
- Neuweg, G. H. (2014). Das Wissens der Wissensvermittler. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 583–614). Waxmann.
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement – Ein Handbuch für Studium und Praxis*. Kohlhammer.
- Reuker, S. (2017a). The noticing of physical education teachers: a comparison of groups with different expertise. *Physical Education and Sport Pedagogy*, 22(2), 150–170. <https://doi.org/10.1080/1748989.2016.1157574>
- Reuker, S. (2017b). The knowledge-based reasoning of physical education teachers: A comparison between groups with different expertise. *European Physical Education Review*, 23(1), 3–24. <https://doi.org/10.1177/1356336X15624245>
- Roose, I., Goossens, M., Vanderlinde, R., Vantieghem, W., & Van Avermaet, P. (2018). Measuring professional vision of inclusive classrooms in secondary education through video-based comparative judgement: An expert study. *Studies in Educational Evaluation*, 56(1), 71–84. <https://doi.org/10.1016/j.stueduc.2017.11.007>
- Römer, J., & Rothland, M. (2015). Klassenführung in der deutschsprachigen Unterrichtsforschung. Ein kritischer Überblick zur Operationalisierung und empirischen Erfassung. *Empirische Pädagogik*, 29(2), pp. 266–287.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771. <https://doi.org/10.3102/0002831214531321>
- Shavelson, R. (2013). On an approach to testing and modeling competence. *Educational Psychologist*, 48(2), 73–86. <https://doi.org/10.1080/00461520.2013.779483>
- Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Söll, W., & Kern, U. (2008). *Sportunterricht - Sport unterrichten: Ein Handbuch für Sportlehrer*. Hofmann.
- Ulferts, H. (2019). *The relevance of general pedagogical knowledge for successful teaching: Systematic review and meta-analysis of the international evidence from primary to tertiary education*. OECD Publishing. <https://doi.org/10.1787/19939019>
- van der Mars, H., Darst, P., Vogler, B., & Cusimano, B. (1994). Active Supervision Patterns of Physical Education Teachers and Their Relationship with Student Behaviors. *Journal of Teaching in Physical Education*, 14(1), 99–112. <https://doi.org/10.1123/jtpe.14.1.99>
- van Driel, S., Wolff, C. E., Crasborn, F., Brand-Gruwel, S., & Jarodzka, H. (2022). A coding scheme to clarify teachers' interactive cognitions in noticed classroom management situations from an actor's perspective. *Teaching and Teacher Education*, 111, 103602. <https://doi.org/https://doi.org/10.1016/j.tate.2021.103602>
- van Es, E.A., & Sherin, M.G. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187–223. <https://doi.org/10.1007/s11618-015-0626-6>
- Walkoe, J., Sherin, M., & Elby, A. (2020). Video tagging as a window into teacher noticing. *Journal of Mathematics Teacher Education*, 23(4), 385–405. <https://doi.org/10.1007/s10857-019-09429-0>
- Weber, K. E., Gold, B., Prilop, C. N., & 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. *Teaching and Teacher Education*, 76, 39–49. <https://doi.org/10.1016/j.tate.2018.08.008>
- Wibowo, J., & Heemsoth, T. (2019). Fachdidaktische Wissen von Sportlehrer*innen testen: Überlegungen zur Inhaltsvalidität. *Zeitschrift für sportpädagogische Forschung*, 7(2), 88–108.

Keywords: Education of physical education teachers; test development; content validation; Delphi study; classroom management

Von Wissen zur Performanz in der Ausbildung von sportunterrichtenden Lehrpersonen: Eine Delphi-Studie und ein Pretest zur inhaltlichen Validierung der Testinstrumente

Zusammenfassung

Die Basis zur Eruierung der Effektivität von Interventionen in der Aus- und Weiterbildung von sportunterrichtenden Lehrpersonen stellen valide, reliable und objektive Instrumente dar, die weitgehend fehlen. Zur Entwicklung derartiger Instrumente ist die Inhaltsvalidität eine notwendige Grundlage. In vorliegender Studie wird dementsprechend die inhaltliche Validierung eines Tests zur Erfassung des (jeweils klassenführungsbezogenen) professionellen Wissens und der Wahrnehmung, Interpretation und Entscheidung (PID) bei (angehenden) sportunterrichtenden Lehrpersonen dargestellt. Es resultiert ein inhaltlich validierter Wissenstest (110 Testitems) sowie ein videobasierter Test zur Erfassung der PID (zehn Vignetten, 217 Items). In einem nächsten Schritt werden die Instrumente empirisch validiert.

Schlagworte: Aus- und Weiterbildung von Sportlehrpersonen; Testentwicklung; Inhaltliche Validierung; Delphi-Studie; Klassenführung

Le processus transformatif du savoir à la performance chez les futur-e-s enseignant-e-s en éducation physique: étude Delphi et pré-test pour la validation du contenu des instruments de test

Résumé

Le fondement de toute l'évaluation de l'efficacité des interventions dans la formation de base des enseignant-e-s en éducation physique (EP) repose sur des instruments, qui à ce jour font largement défaut. Pour développer de tels instruments, la validité du contenu est une base. Cette étude présente donc la validation du contenu d'un test destiné à évaluer le savoir professionnel (liée à la gestion de la direction de classe) et la perception, l'interprétation et la décision (PID) chez les (futur-e-s) enseignant-e-s en EP. Le résultat de cette démarche empirique est un test de connaissances validé quant à son contenu (110 items) ainsi qu'un test basé sur de la vidéo pour saisir les éléments du PID (dix vignettes, 217 items). Dans une prochaine étape, les instruments seront validés empiriquement.

Mots-clés: Formation de base et formation continue des enseignant-e-s en éducation physique; développement de tests; validation du contenu; étude Delphi; gestion de classe

Sul processo di trasformazione dalla conoscenza alla performance nei futuri insegnanti: Studio Delphi e pretest per la validazione del contenuto degli strumenti di test

Riassunto

La base per valutare l'efficacia degli interventi nella formazione e nello sviluppo professionale continuo degli insegnanti di educazione fisica (EF) prevede strumenti validi, affidabili e oggettivi, che risultano in gran parte carenti. Per sviluppare tali strumenti, la validità del contenuto è indispensabile. Questo saggio presenta la validazione del contenuto di un test per la valutazione delle conoscenze professionali (relative alla gestione dell'aula) e della percezione, interpretazione e decisione (PID) dei futuri insegnanti di educazione fisica. Il risultato è un test di conoscenza validato nei contenuti (110 item) e un test basato su video per la valutazione della PID (dieci vignette, 217 item). In una fase successiva, gli strumenti saranno validati empiricamente.

Parole chiave: Formazione di base e continua degli insegnanti di educazione fisica; sviluppo di test; validazione dei contenuti; studio Delphi; gestione della classe

Matthias Baumgartner, Dr. phil., is the head of the Institute of Physical Education, Sport and Health (IPESH) at the University of Teacher Education St. Gallen. His research focuses on the areas of competence-oriented physical education teacher education research, teaching research and research of higher education.
PH St. Gallen, Notkerstrasse 27, CH-9000 St. Gallen
E-mail: matthias.baumgartner@phsg.ch

Eric Jeisy, Dr. phil., works as a senior researcher at the University of Teacher Education St. Gallen. His research and development focuses on learning and teaching and digital transformation in the subject area of movement and sport, sport-related teacher education and professional research.
PH St. Gallen, Notkerstrasse 27, CH-9000 St. Gallen
E-mail: eric.jeisy@phsg.ch

Clemens Berthold, Mag., is currently employed as a doctoral student at the University of Teacher Education St. Gallen. He studied physical education and physics for teacher education at the university of Vienna. He was student assistant at the institute of physical education and later worked as a teacher.
PH St. Gallen, Notkerstrasse 27, CH-9000 St. Gallen
E-mail: clemens.berthold@phsg.ch