Journal for Educational Research Online
Journal für Bildungsforschung Online
https://doi.org/10.31244/jero.2023.01.03
Volume 15 (2023), No. 1, 51–73

Wiebke Langer, Julia Bruns, & Jan Erhorn

Development and Validation of a Video-Based Test Instrument to Measure *Noticing With Regard to Providing Individual Learning Support* in Inclusive Physical Education

Abstract

Noticing is considered to be of crucial importance for professional action of physical education (PE) teachers especially in inclusive PE. In order to ensure an appropriate consideration of individual learning conditions and processes in the planning and implementation of lessons, teachers must selectively focus their attention on events in a complex and interactive situation in PE that are relevant for the individual learning support of the students and interpret them on a theoretical basis. Noticing, which includes selective attention and knowledge-based reasoning, is regarded as a central prerequisite for enabling the (PE) teacher to provide all students individual learning support in PE. It has not yet been investigated how well physical education teacher education (PETE) programs succeed in promoting the competence facet of noticing in the area of individual learning support in inclusive PE. This is also due to the fact that there are not yet any instruments for the valid psychometric measurement of the construct. In the context of this paper, the development and validation of a standardized, video vignette-based test instrument for the measurement of noticing of prospective PE teachers with regard to providing individual learning support is presented. The validation focuses on the content of the test and the internal structure of the instrument as well as on correlations with conceptually related constructs. The findings to date indicate

Dr. Wiebke Langer (corresponding author), ORCID: 0000-0002-0877-7469, University of Hamburg, Institute of Human Movement Science, Turmweg 2, 20148 Hamburg, Germany, and Institute of Sports and Movement Science, Osnabrück University, Germany
email: wiebke.langer@uni-hamburg.de

Jun. Prof. Dr. Julia Bruns, ORCID: 0000-0002-6604-5864, Faculty of Computer Science, Electrical Engineering and Mathematics, Didactics of Mathematics, Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany
email: julia.bruns@uni-paderborn.de

Prof. Dr. Jan Erhorn, ORCID: 0000-0002-0684-6707, Institute of Sports and Movement Science, Osnabrück University, Jahnstr. 75, 49080 Osnabrück, Germany
email: jan.erhorn@uni-osnabrueck.de

that the test instrument ViProQiS_IF was able to provide a standardized measurement of noticing with regard to providing individual learning support in inclusive PE.

Keywords
physical education teacher education (PETE), competence modeling and measurement, inclusive physical education, video-based items, noticing

Entwicklung und Validierung eines videobasierten Testinstruments zur Erfassung des *Noticing mit dem Fokus auf individuelle Förderung* im inklusiven Sportunterricht

Zusammenfassung
Noticing besitzt eine hohe Bedeutung für das professionelle Handeln von Sportlehrkräften, insbesondere für den inklusiven Sportunterricht. Um eine angemessene Berücksichtigung individueller Lernvoraussetzungen und -prozesse bei der Planung und Durchführung von Unterricht zu gewährleisten, müssen die Sportlehrkräfte ihre Aufmerksamkeit gezielt auf die für die individuelle Förderung der Schüler:innen relevanten Ereignisse im zum Teil unübersichtlichen interaktiven Geschehen in der Sporthalle richten und diese auf theoretischer Grundlage interpretieren. Noticing, welches die selektive Aufmerksamkeitslenkung und das theoriegeleitete Deuten umfasst, bildet somit eine wichtige Voraussetzung, um allen Schüler:innen eine individuelle Förderung im inklusiven Sportunterricht zu ermöglichen. Die Fähigkeiten angehender Sportlehrkräfte im Bereich Noticing insbesondere in inklusiven Settings sind bislang kaum untersucht worden. Diese Forschungslücke kann u. a. auf fehlende Forschungsinstrumente zur standardisierten Erfassung dieses Konstrukts zurückgeführt werden. Im Rahmen des Beitrags wird daher die Entwicklung und Validierung eines standardisierten, Videovignetten-basierten Testinstruments zur Messung des Noticing von angehenden Sportlehrkräften im Hinblick auf individuelle Förderung im inklusiven Sportunterricht vorgestellt. Im Fokus der Validierung stehen der Testinhalt, die interne Struktur des Instruments sowie Zusammenhänge mit verwandten inhaltlichen Konstrukten. Die Ergebnisse aus drei aufeinander aufbauenden Validierungsstudien geben Hinweise auf die Reliabilität und Validität der Testwertinterpretation des Instruments ViProQiS_IF im Sinne des Noticing im Hinblick auf individuelle Förderung im inklusiven Sportunterricht.

Schlagworte
Sportlehrkräftebildung, Kompetenzmodellierung und -messung, inklusiver Sportunterricht, Videovignetten-basierte Items, Noticing

1. Introduction

Providing individual learning support is an important demand of (inclusive) PE in addressing heterogeneity in an appropriate way (Erhorn, Langer, & Möller, 2020; Neuber & Pfitzner, 2012). In doing so, it should consider the different learning conditions and the associated individual potentials of the students. The fit between a learning task and the students' individual learning conditions is seen as a central requirement for providing individual learning support. In order to ensure an appropriate consideration of individual learning conditions and processes in the planning and implementation of lessons, teachers must selectively focus their attention on events in a complex classroom situation that are relevant to identify the fit/misfit between a learning task and individual learning conditions at a motor, cognitive, and motivational-affective level in specific contexts and interpret them on a theoretical basis (Krammer et al., 2016; Reuker, 2018). Van Es and Sherin (2002) refer to this selective attention and knowledge-based reasoning as the situated competence facet of *noticing*. They emphasize that prospective teachers should be taught skills in noticing, as these processes already represent a key element of coping with demands and thus offer an opportunity to establish a link between theoretical training content and prospective teaching practice (van Es & Sherin, 2008). It has not yet been investigated how well physical education teacher education (PETE) programs succeed in promoting the competence facet of noticing in the area of individual learning support in inclusive PE. This is also due to the fact that there are not yet any instruments for the valid psychometric measurement of the construct (Erhorn, Moeller, & Langer, 2020; Reuker et al., 2016). This paper addresses this research gap by presenting the development and validation of the video vignette-based instrument ViProQiS_IF, which is designed to measure the construct of *noticing with regard to providing individual learning support*.

2. Theoretical Framework and State of Research

2.1 Individual Learning Support as an Aim of Inclusive PE

Individual learning support has become a recognized quality criterion for educational programs and good teaching (Corno, 2008; Ní Bhroin & King, 2020; Pfitzner & Neuber, 2012; United Nations Educational, Scientific, and Cultural Organization [UNESCO], 1994). Even though the increased individualization of teaching and learning is not a new requirement and is already embodied in school and teacher training laws, it has become more important and topical due to the global trend towards inclusive schooling (Dumont, 2019; Neuber & Pfitzner, 2012; Ní Bhroin & King, 2020; Sawalies et al., 2013). There is also widespread consensus in the sport-didactic professional discourse about the relevance of individual learning support in PE under conditions of heterogeneity (Block et al., 2017; Lieberman & Houston-Wilson, 2018; Neuber & Pfitzner, 2012). Overall, however, it is apparent that indi-

vidual learning support should not be understood as a uniform concept, but nevertheless as a general demand on (inclusive) teaching. Currently, for example, the approaches of adaptive teaching, scaffolding, and formative assessment, or assessment for learning are often discussed as variants of individual learning support in regular teaching practices (Dumont, 2019). Although the concepts differ, they agree on the guiding idea of enabling comprehensive development of the potential of all students and taking into account their individually different learning conditions. In doing so, adaptive teaching can be understood as a basic orientation to take into account interindividual differences of the students in a didactically appropriate way (Dumont, 2019; Hardy, 2017). For individual supportive teaching in (inclusive) PE, this means that every student should be given the opportunity to "comprehensively develop their motor, intellectual, emotional, and social potential" and receive optimal support through appropriate methods (Neuber & Pfitzner, 2012). This requires adaptive teacher action that aims at a fit between the learning tasks and the individual learning conditions of the students (Corno, 2008; Hardy, 2017). In situations of (inclusive) PE, this fit between learning conditions and the requirements of learning tasks must be established particularly at the motor level, at the cognitive level, and at the motivational-affective level. The motor learning conditions include aspects like constitution, physical condition, and coordination skills, and the basic cognitive skills include understanding and assessment in particular, as well as (prior) knowledge with regard to tasks, rules of the game, and competition (Giese & Weigelt, 2017; Lieberman, 2017; Neuber & Pfitzner, 2012). Motivational-affective characteristics refer to different processes of self-related emotional experience in the social learning context (Dresel et al., 2013) and are manifested, for example, in different expressions of the domain-specific self-concept of ability, goal orientation, or intrinsic motivation (Pfitzner & Neuber, 2012). While the motor and cognitive learning conditions are thus more specifically directed towards coping with learning tasks, motivational-affective characteristics have a particular effect at the stages of intention formation and maintenance of the learning act (Wolters et al., 2009).

According to Vygotskij's (1963) concept of zones of proximal development, there is a fit between individual learning conditions and learning tasks performed at the motor, cognitive, and motivational-affective levels when the respective mastery of the requirements is within reach for the learner with the help of external support and thus lies within the zone between his or her current and potential level of development. A problem of fit, on the other hand, is characterized by the fact that the learner is permanently underchallenged or overchallenged. The central challenge for the (PE) teacher is therefore to provide situational and individualized support for each learner in his or her learning and developmental area and thus to ensure the optimal degree of motor, cognitive, and motivational-affective activation (Dumont, 2019). In line with the principle of *scaffolding*, the task is to first build up external learning support and then slowly withdraw it, transferring the responsibility for the learning process gradually to the learner (van de Pol et al., 2010). Such adaptations of the learning process can be made at two levels according to the

adaptive teaching approach (Corno, 1995, 2008; Corno & Snow, 1986): On the one hand, one can make adaptations involving the long-term and overarching design of a differentiated teaching-learning arrangement at the classroom level, so-called *macro-adaptations*. On the other hand, short-term adaptations can be made during the teaching act and the individual learning processes of students, so-called *micro-adaptations* (Corno, 2008; Corno & Snow, 1986). In order to ensure a continuous fit in the learning process, the teaching must be continuously modified so that it corresponds to the further development and change of the individual learning conditions of the students (Corno, 2008; Lieberman, 2017). This requires an accompanying monitoring and assessment of the learning processes and an adaptive use of these indications for individual learning support. Such feedback accompanying the learning process and learning-conducive assessment on the basis of defined learning goals are the core elements of *assessment for learning* (e.g., Black & Wiliam, 2009; Prengel et al., 2009).

The essential precondition for the continuous provision of such opportunities for individual learning support is that the (PE) teacher is able to selectively direct his or her attention to events that are relevant to learning as well as those that hinder it and to interpret them on a theoretical basis.

2.2 Noticing as a Situated Facet of Competence

The task of continuously optimizing individual learning processes in inclusive PE classes places high situational demands on the (PE) teacher. In order to flexibly adapt their lessons to the individual learning conditions of the students and design suitable teaching and learning opportunities, the (PE) teachers must specifically turn their attention to the interplay between learning tasks and individual learning and acquisition processes and thus to the events that foster learning and those that hinder it (Sherin et al., 2011; van Es, 2011; van Es & Sherin, 2002), identify learning situations in a differentiated way, analyze them, and ultimately assess the learning process with regard to the fit between the learning opportunities and individual learning conditions (Erhorn, Langer, & Möller, 2020; Reuker, 2018). The concept of noticing (van Es & Sherin, 2002) can be used to determine the competence facets required for this. The focus here is on the situated processes of *selective attention* and *knowledge-based reasoning* of perceived events (van Es & Sherin, 2002). These are regarded as central prerequisites for enabling the (PE) teacher to handle teaching requirements professionally in different situations (Kramer et al., 2017; Krammer et al., 2016; Reuker, 2018; van Es & Sherin, 2002, 2008). These situated processes of perception are relevant to ensure that a teacher can distinguish important from unimportant events in complex classroom situations and use their attention and skills where they are needed to act adaptively (Kramer et al., 2017; Krammer et al., 2016; Reuker, 2018; van Es & Sherin, 2002, 2008).

Findings from research on expertise (Berliner, 2001) demonstrate that the noticing of teaching by experts and novices exhibits qualitative differences with regard to

the selective directing of attention, the differentiated description of situation-specific relevant teaching situations, and their explanation or interpretation on the basis of theoretical knowledge. Prospective and novice teachers in particular, who often cannot identify events that are relevant to learning (Star & Strickland, 2008) or distinguish important from unimportant events (Berliner, 2001), are also not able to interpret relevant situations and link them to theories and background knowledge. Novices thus predominantly remain at the level of undifferentiated or small-scale descriptions of teaching events and have a tendency to overgeneralize evaluations and assessments on the basis of surface characteristics. Sabers and colleagues (1991), for example, found in their expert-novice study that novices often focus on the behavior of students, especially when they do not behave according to expectations. In doing so, these prospective teachers neither make assumptions about the causes that led to this behavior, nor do they formulate possibilities for resolving the challenging situation. Experts, on the other hand, are able to perceive more complex teaching situations and systematic problems pertaining to a complex learning process on the basis of their specific knowledge structures. They classify problems at a high level of abstraction, can reduce them to the essentials and consolidate them into units of meaning, and thus arrive more quickly and in a more differentiated manner at diagnoses, explanations, and classifications of structures or contexts that are relevant to learning as well as at solutions to problems (Berliner, 2001; Sabers et al., 1991; Seidel & Prenzel, 2007; Seidel & Stürmer, 2014). Overall, the expert-novice studies point to an increase in the quality of teaching perception depending on (professional) experience and, at the same time, pronounced professional knowledge (Berliner, 2001; Seidel & Prenzel, 2007; Seidel & Stürmer, 2014).

The fact that PE differs in many components from other school subjects (e.g., focus on movement and related motor and social processes) results in subject-specific requirements regarding the noticing of PE teachers. Content-specific knowledge and pedagogical content knowledge are assumed to be required qualifications for noticing the more subject-specific characteristics focused on here, such as the fit between learning tasks and individual prerequisites at a motor, cognitive, and motivational-affective level (Meschede et al., 2015). Against this background, noticing is considered to have domain-specific importance for individual learning support in inclusive PE (Blomberg et al., 2011; Meschede et al., 2015; Steffensky et al., 2015).

2.3 Approaches to Measuring Situated Competences

The measurement of situated competence facets like noticing requires methods that enable a contextualization that is authentic and relevant to teaching requirements (Hoth, 2016; Knievel et al., 2015; Lindmeier, 2013; Seidel & Thiel, 2017; Wagner & Ehlert, 2018). The use of vignettes in the form of texts, comics, photos, or (scripted) video sequences has become the established method for the realistic depiction of concrete teaching situations as well as individual learning and acquisition processes (e.g., Bruns et al., 2020; Friesen et al., 2018; Hoth, 2016; Kersting, 2008;

Kramer et al., 2017; Lindmeier, 2013; Wagner & Ehlert, 2018). Such vignettes serve as stimuli that can be used to measure processes of selective attention and theory-based interpretation of relevant teaching situations as well as decision-making in open task formats or closed rating items (Bruns et al., 2020; Kramer et al., 2017; Seidel & Thiel, 2017). Recent studies favor the use of real teaching sequences to test situated competence facets in prospective teachers, especially processes of selective attention and knowledge-based reasoning (e.g., Bruns et al., 2020; Kersting, 2008; Seidel et al., 2010). The video sequences make real demands on the teacher and enable a holistic representation of complex interactions between students and teachers (e.g., König, 2015; Kramer et al., 2017; Seidel & Thiel, 2017). In addition to the presentation of the stimuli, there is also a discussion in the literature of different task formats, that is, open and closed, and of their effects on the assessment of situated competence facets. Open tasks indeed require a more elaborate evaluation process, but they are seen as offering advantages mainly in terms of selective processes of identification. The focus is less directed or anticipated than in the case of standardized rating items, thus enabling a fundamental intuitiveness of perception, expressed interpretations, and options for action (e.g., Lindmeier, 2013; Meschede et al., 2015).

Although previous studies have measured professional perception specific to PE (Reuker, 2017a, 2017b, 2018), also in the context of inclusive PE (Reuker & Rischke, 2017), there are not yet any valid test methods for the standardized measurement of situated competence facets in the disciplinary discourse.

The lack of appropriate methods that are at the same time content-specific and sufficiently sensitive led us to develop the video vignette-based instrument. A separate test scale is being constructed for each of the three central demands of inclusive PE: Recognition, Individual Learning Support, and Participation (Langer, Bruns, & Erhorn, 2022; Langer, Bruns, & Erhorn [in press]). In these studies, we present the instrument ViProQiS_IF with a focus on the construct of *noticing with regard to providing individual learning support*.

3. Test Construction

The test construction of the ViProQiS_IF instrument is based on the rational method (Eid & Schmidt, 2014, p. 57). It followed the measurement approach of construct modeling (Wilson, 2005) with the four building blocks construct map, item design, outcome space, and measurement model.

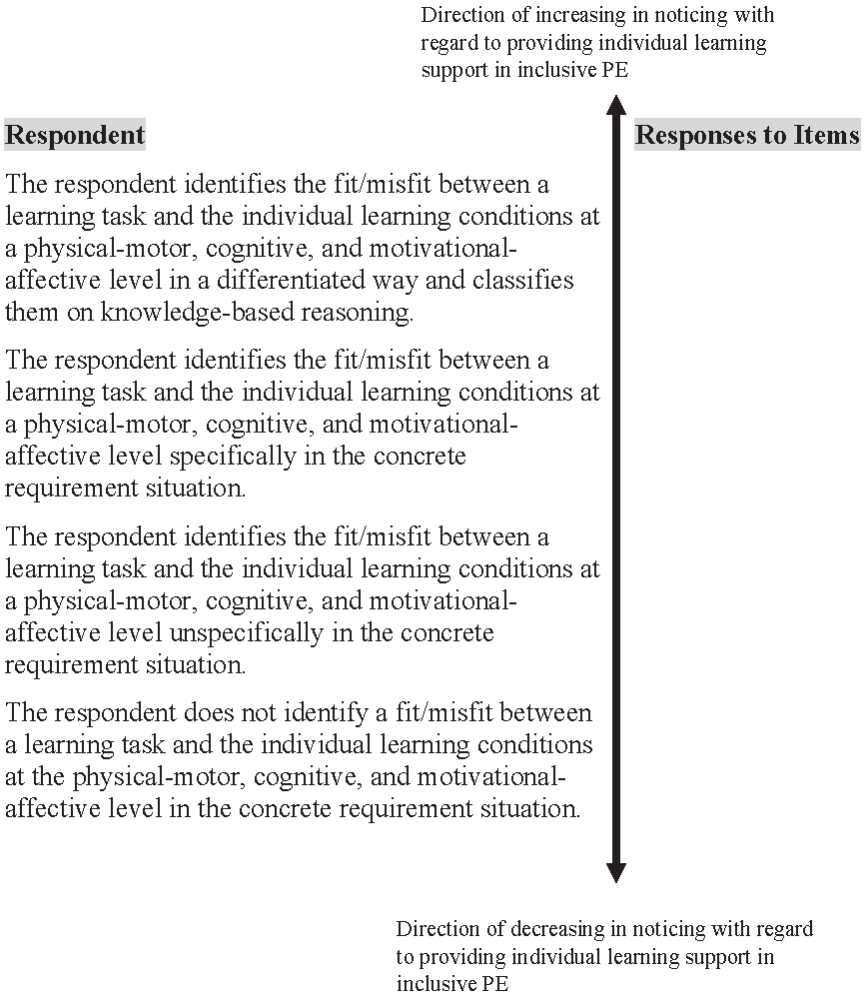
3.1 Definition of the Theoretical Construct

According to Wilson (2005), the basic preconditions of construct modeling are a coherent theoretical definition of the target construct and the plausible assumption that the ability being measured can be mapped on an underlying continuum from

high to low expression. The boundaries or extreme expressions of the construct are defined and the transformation of the theoretical construct into a unidimensional latent variable prepared through the development of a construct map (Wilson, 2005, pp. 25–28).

In order to model noticing with regard to providing individual learning support in inclusive PE, we included the fit between learning conditions and the demands of learning tasks at the motor, cognitive, and motivational-affective levels in situations of inclusive PE (see Section 2.1). In the process, we took into account that selective attention is a cognitive process and therefore cannot be measured directly but rather has to be mapped by an observable process, in this case identifying. The creation

Figure 1: Construct Map for the Construct Noticing With Regard to Providing Individual Learning Support, Adapted From Wilson (2005)



of different levels in the corresponding construct map (see Figure 1) follows the assumption that persons with a low level of the competence facet remain at the level of describing in an unfocused observation, whereas persons with a high level of the competence facet structure teaching processes into higher-level sense units, identify more complex connections, and consequently formulate adequate knowledge-based explanations and evaluations (Berliner, 2001; Seidel & Prenzel, 2007).

We therefore defined noticing with regard to providing individual learning support in inclusive PE as the ability of (prospective) PE teachers to identify the fit/misfit between a learning task and individual learning conditions at a motor, cognitive, and motivational-affective level in specific contexts and to interpret them on a theoretical basis.

3.2 Presentation of the Test Instrument

The development of test items on the model provided by Wilson (2005) aims at appropriately operationalizing the theoretical construct presented in Section 3.1 in its gradation. Video vignette-based standardized test instruments are increasingly used for the situated assessment of noticing (e.g., Kramer et al., 2017; Seidel & Thiel, 2017). In contrast to procedures with text vignettes, they offer the advantage of the necessary situation-specific contextualization (Blömeke et al., 2015; see also Section 2.2). On the basis of these considerations, we used video vignettes as test stimuli for activating noticing in a targeted manner, that is, with regard to the aim of providing individual learning support.

The video vignettes are 1- to 2-minute video sequences from a video data corpus of 104 hours of videotaped PE lessons in inclusive settings[1] (Erhorn & Langer, 2022; Erhorn, Langer, & Möller, 2020). For this purpose, we filmed 14 classes at six schools for a complete lesson using two cameras. The non-scripted video sequences allow for an authentic depiction of everyday school life. In the absence of clear research-based quality criteria for determining video sequences of inclusive PE, the selection of videotaped case studies is based on expert judgement (see also Section 4.1). On this basis, we developed eight items, each consisting of a prompt and a video vignette, to assess noticing with regard to providing individual learning support.

The item prompt provides a brief classification of the situation and asks the test subject to analyze the situation with regard to providing individual learning support. We chose a maximum open-response format without time pressure for the survey in order to allow for a breadth of content in the intuitiveness of noticing in the teaching situation shown in the respective video sequence (Eid & Schmidt, 2014, pp. 98–99). Table 1 shows three video vignettes as examples and the fit/misfit between a learning task and individual learning conditions at a motor, a cognitive,

1 The video recordings of inclusive physical education were taken at schools in Hamburg. In Hamburg, all students have a legal right to access a public school.

and a motivational-affective level in specific contexts that could obviously be identified on the basis of the expert judgement.[2]

Table 1: Description of Three Exemplary Video Vignettes

Video vignette	Description of the video vignette	The fit between a learning task and individual learning conditions
1: Jump and crawl obstacle course	The video vignette shows a PE lesson sequence at primary school. It is a lesson of the teaching unit on the topic "gym". A fitness course is set up. In the foreground, a boy with a black T-shirt can be seen who has great motor difficulties in overcoming an obstacle (a box).	This obstacle station in the fitness course therefore presents a problem of fit for this student in his individual motor conditions.
3: Pylons drop	The video vignette shows a PE lesson sequence during a game to practice throwing at primary school. Two teams play against each other. Each team has a box in the playing field with four pylons on it. The aim is to throw the pylons away from the opposing team. To do this, two balls are brought into play. In the foreground is a girl with trisomy 21 who has great difficulty participating in the game.	In this complex game situation, it is obvious that the girl is motivated to participate in the game. However, despite individual verbal explanations and demonstrations by the teacher and active support provided by other students, she does not succeed in understanding and following the gameplay in a self-determined way. Against this background, it can be analyzed in this situation that the game presents a problem of fit for this girl in her individual cognitive conditions.
4: Group warm-up exercise	The video vignette shows a PE lesson sequence during a group warm-up exercise at secondary school. Three students can be seen fooling around and not following the teacher's instructions.	In this situation, it is obvious that the three students are fooling around and do not perform the exercises in a focused and correct manner, although they would be able to do so. The exercises do not interest them while being underchallenged. The group warm-up exercise presents a problem of fit for these three students in their individual motivational-affective conditions.

As a means of drawing conclusions about the expression of this construct in the test subjects from their responses to the items, the responses are first categorized on the basis of a coding manual and then processed for subsequent modeling, that is, transferred into scores on the continuum of the latent construct (Wilson, 2005, p. 69). The coding of the eight open-ended items follows a coding manual developed in an iterative process (Eid & Schmidt, 2014, p. 244). The scoring of the response categories involves differentiating the response qualities into three levels: no fit identified (Score 0), fit non-specifically to specifically identified (Score 1), fit specifically identified in their more complex contexts and classified on the basis of knowledge (Score 2). In this way, we developed a coding instruction consisting of a coding rule and an anchor example for each item (see also Figure 2). To verify

2 The items and the full test scale are available and can be requested from the corresponding author.

the coding instructions, we had 10 % of the test data double-coded by two trained raters. The interrater reliability, measured by Cohen's kappa, shows a very good average agreement: κ = .82 (range .80–.97; Wirtz & Caspar, 2002, p. 59).

Figure 2: Exemplary Excerpt From the Coding Manual for Item 1 "Jump and Crawl Obstacle Course"

Direction of increasing in noticing with regard to providing individual learning support in inclusive PE

Respondent	Responses to Item 1 (model answers)	Explanation	Coding
The respondent identifies the fit/misfit between a learning task and the individual learning conditions at a physical-motor, cognitive, and motivational-affective level in a differentiated way and classifies them on knowledge-based reasoning.	"[…] The teacher tries to provide situational and individualized support for the boy in the black shirt in his learning and developmental area: The teacher knows that the boy has movement difficulties and also knows that encouragement leads him to be more confident in his movement actions. She knows that a short-term modification of the movement task (reducing the level of motor requirements: no longer jumping, but stepping over it) leads to the boy being able to manage the task and does this in terms of individual ability […]." (xZStO)	Problem of fit between the learning tasks and the individual learning conditions at a physical-motor level is recognized and classified on knowledge-based reasoning.	2
The respondent identifies the fit/misfit between a learning task and the individual learning conditions at a physical-motor, cognitive, and motivational-affective level specifically in the concrete requirement situation.	"The task 'jumping over the box' overchallenges the boy at a physical-motor level. The teacher tries to adapt the task to him by helping him and simplifying the task at its motor requirement level." (iyZLs)	Problem of fit between the learning tasks and the individual learning conditions is recognized and specifically explained.	1
The respondent identifies the fit/misfit between a learning task and the individual learning conditions at a physical-motor, cognitive, and motivational-affective level unspecifically in the concrete requirement situation.	"The teacher helps a boy who has problems jumping over a box. The teacher tries to help him." (N6Yta)	Problem of fit between the learning tasks and the individual learning conditions is recognized but only unspecifically explained.	
The respondent does not identify a fit/misfit between a learning task and the individual learning conditions at the physical-motor, cognitive, and motivational-affective level in the concrete requirement situation.	"The teacher asks the student to jump over the first obstacle and the student falls down. She says it is not a problem and next time he crosses the obstacle he should not jump but instead step over it […]." (2FhTK)	No mention of a specific fitting problem, only a description of the situation.	0

Direction of decreasing in noticing with regard to providing individual learning support in inclusive PE

4. Evidence on the Validity of the Test Score Interpretation

Developing a new instrument for measuring teachers' noticing requires paying special attention to the validity of the test score interpretation. "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of the test" (American Educational Research Association [AERA] et al., 2014, p. 11). The ViProQiS_IF test instrument aims at measuring how well (preservice) PE teachers notice students' need for individual learning support. The instrument should therefore measure whether participants recognize situations in which a learning task does not fit students' learning conditions at a motor, cognitive, and motivational-affective level. The instrument was developed for research purposes, such as examining effects of professional development, rather than for individual assessments. To validate whether the test score can be interpreted as a measure of how well PE teachers notice their students' need for individual learning support, we focused on four aspects of the instrument: test content, response processes, internal structure, and convergent validity (AERA et al., 2014).

4.1 Evidence Regarding the Test Content

Sources of evidence regarding the content of the test result partly from the theory-based development of the video-based items. For this purpose, we adapted van Es and Sherin's (2002) concept of noticing and applied it to the aim of providing individual learning support in (inclusive) PE.

For the item development, eight researchers with a proven research focus in the field of inclusive PE and expertise in video-based instructional analysis assessed 20 preselected video sequences for their suitability as stimuli. These sequences are different categorically evaluated classroom situations of varying complexity (Erhorn, Langer, & Möller, 2020; Oser, 2013) that take into account the fit/misfit between the demands of a learning task and individual learning conditions at the motor, cognitive, and motivational-affective levels. In addition to ensuring their authentic representativeness, we assessed the sequences independently with regard to whether they were appropriate for the content or obviously fit the theoretical construct as well as their location on the scale of the construct map or their difficulty. We then discussed and evaluated them within the framework of a consensus validation (Wilson, 2005, pp. 59–60). The experts' comments during the panel discussion were digitally recorded and then qualitatively evaluated. On this basis, we selected eight suitable video sequences, three of which depicted the fit between a learning task and individual learning conditions at a motor and cognitive level and two of which depicted this fit at a motivational-affective level. These video vignettes were then provided with the corresponding prompts and presented to the experts again for a final assessment as a means of checking the quality of the items. Overall, this

process ensured that the theoretical construct was appropriately operationalized in its continuous expression and that the selected video vignette-based items represented authentic situations for individual learning support in inclusive PE. In addition, these items of varying complexity differ in whether the focus of attention is directed (Items 1, 2, 3, 5, 8) or non-directed (Items 4, 6, 7). To evaluate the validity of teachers' response processes, in a next step, we conducted a cognitive lab study (Collins, 2003) with $N = 8$ preservice PE teachers using the thinking aloud approach (Ericsson & Simon, 1993). The cognitive lab study aimed at examining the extent to which the items actually trigger cognitive processes relevant to our construct in the respondents (Schnell, 2016). With the items, it was possible to provoke procedures of noticing the fit/misfit between a learning task and the individual learning conditions of a single student or of different students at a motor, cognitive, and motivational-affective level in a complex PE teaching situation in a context- or situation-specific manner. No unwanted comprehension problems occurred in the prompts. All in all, the items succeeded in activating noticing with regard to providing the students individual learning support at different levels and in solving the tasks on the basis of the stimuli. The results show that correct answers were based on the intended cognitive and situation-specific processes of noticing, whereas incorrect answers could be attributed to inappropriate cognitive and situation-specific processes of noticing. Inappropriate cognitive and situation-specific processes of noticing were especially observed in responding to the items with high complexity, that is, with non-directed focus of attention, if the item was about individual learning support of different students. For example, in Item 4, one of the responses from a preservice PE teacher focused on a child moving in the back who has forgotten his sports clothes and tries to attract the teacher's attention by commenting aloud on the other classmates' practice, instead of analyzing the group warm-up and the misfit visible there between the exercises and the individual learning conditions.

The preservice teachers' responses showed fewer differences in terms of focus on the intended stimulus in each case than in terms of depth of perception and, as a result, knowledge-based inference. For example, in Item 1, all participants identified that there was a misfit issue between a learning task and a student's individual learning conditions. Via the specific perception of the obvious fit issue, two preservice teachers were able to relate it in a differentiated way with regard to motor factors of the student's individual learning conditions and interpret the correct descriptive features of micro-adaptations that occurred and their desired macro-adaptations, as the following example shows:

> At the teacher's start signal, the student in the black shirt dynamically starts jumping over the box as required and abruptly stops directly in front of the obstacle. The student is obviously fully motivated and willing to tackle the obstacle station on his own, but he does not succeed. This obstacle station in the fitness course therefore clearly presents a problem of fit for this student in the requirements on the motor level. The teacher notices this and approaches the student to help him. When it becomes clear that this station is causing him motor difficulty,

> she gives him the opportunity to step over the box, which makes the movement task easier. Although a problem of fit arises, her assistance enables him to complete the task. She might have provided the station with a differentiated set-up or task beforehand to address the heterogeneity of the students. (Interview with Preservice PE Teacher 3)

4.2 Validity Evidence Concerning the Internal Structure of ViProQiS_IF

To provide the validity evidence concerning the internal structure, we piloted the ViProQiS_IF instrument with $N = 261$ preservice PE teachers, all studying at one university in Germany. On average, the participants were 27.43 years old and in the fourth semester of their studies; $n = 102$ of the respondents identified as male, $n = 152$ as female, and $n = 7$ respondents gave no information. In addition, we asked the participants the number of hours the topic of inclusion in general ($M = 19.36$, $SD = 34.61$, Min. = 0, Max. = 200) and the content-specific topic of inclusive PE had already been treated in the context of their studies ($M = 7.80$, $SD = 18.10$, Min. = 0, Max. = 170).

Data were collected in a group setting in different seminar groups of the Bachelor's and in an introductory lecture of the first semester within 2 months. Therefore, all questions and test items were integrated into an online survey. The video vignettes were presented to the whole group. Afterwards, the students answered the items via the online survey. Data collection was conducted by specially trained test assistants.

We evaluated the internal structure of the test instrument on the basis of a confirmative factor analysis conducted with the R software package lavaan (Rosseel & Jorgensen, 2019). As we assume that the construct is unidimensional (see Section 3.1), we tested a one-factor model against a three-factor model. The three-factor model takes into account three levels that directly influence the task-related learning process (motor level, cognitive level, and motivational-affective level). Next, we examined the quality of the items based on the item response theory using a partial-credit (PC) model (Masters, 1982). The items were evaluated on the basis of their infit, and items with an infit between 0.75 and 1.33 (Wilson, 2005) proved to be acceptable. The analysis was conducted with the R package tam (Rosseel & Jorgensen, 2019).

In a first step, we analyzed the variance of the items. All items showed at least one false, one partly correct, and one correct answer. Therefore, we included all items in the confirmatory factor analysis. We compared a unidimensional model to a three-dimensional model. In the three-dimensional model, the first factor includes only items indicating a fit/misfit between a learning task and the individual learning conditions of a single student or of different students at the motor level (Items 1, 5, 7), the second factor includes only items indicating a fit/misfit between a learning task and the individual learning conditions of a single student or of dif-

ferent students at the cognitive level (Items 2, 3, 8), and the third factor includes only items indicating a fit/misfit between a learning task and the individual learning conditions of a single student or of different students at the motivational-affective level (Items 4, 6).

Table 2: Model Fit Indices of the Compared Models

	df	χ^2	*p*	RMSEA	CFI	SRMR	AIC	BIC	adjBIC
1-dim model	20	10.304	.962	.000 [.000; .000]	1.000	.018	2986.374	3071.552	2995.465
3-dim model	17	6.559	.989	000 [.000; .000]	1.000	.015	2988.630	3084.455	2998.857

Note. *df* = degrees of freedom; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayes information criterion; adjBIC = sample size-adjusted BIC information criterion.

Model fit indices of the two models are compared in Table 2. All indices indicate a good fit of both models to our data, and there are only marginal differences between the AIC, BIC, and adjBIC scores of the two models. A χ^2 test showed no differences between the two models, $\chi^2(3) = 3.745$, $p = .290$. As the unidimensional model is also supported by our theoretical perspective, we examined the unidimensional model in more detail. A closer look at the unidimensional model shows that all items loaded significantly on the common factor (standardized λ between .58 and .71).

In the next step, we examined whether the eight items fit the PC Rasch model. All items showed a good fit, with infit scores between 0.71 and 1.33. However, the scale as a whole showed low reliability (reliabilityWLE = .697; reliabilityEAP = .775). This might have been due to the difficulty of the items: Thresholds for Category 2 were comparably high for Items 4, 5, 6, 7, and 8 (≥ 3.94).

4.3 Convergent Validity Evidence

In addition to recording the test score of the ViProQiS_IF instrument, we collected information on learning opportunities in the areas of inclusion in general and inclusive sports didactics in particular to investigate the correlations of the test score with conceptually related constructs. We expected the test scores of the instrument to correlate positively with the learning opportunities in both areas. Due to the assumed content-specificity of noticing with regard to providing individual learning support (see Section 2.2), there should be stronger correlations between the test score and the subject-specific learning opportunities than between the test score and the general learning opportunities for inclusion. With regard to theoretical experience with inclusive education, we asked the participants the number of hours the topic of inclusive education in general and inclusive PE had already been treat-

ed in the context of their studies at university. In relation to the test score of our instrument ViProQiS_IF, we found a significant correlation between test scores and theoretical experience with inclusive education in general ($n = 253$, $r = .233$, $p < .001$) as well as with inclusive PE ($n = 253$, $r = .166$, $p = .008$), both with a small effect.

5. Discussion of the Evidence of Validity

The situated competence facet of noticing is considered highly relevant for the professional activity of (prospective) (PE) teachers in inclusive settings. In addition, noticing the fit/misfit between a learning task and individual learning conditions at a motor, cognitive, and motivational-affective level was found to be particularly important for the professional management of specific classroom situations in inclusive PE. Against this background, the aim of this paper was the theory-based development and validation of a test instrument that measures noticing with regard to providing individual learning support in inclusive PE. The final instrument, ViProQiS_IF, consists of eight video-based items depicting different situations in inclusive PE in which the fit between the demands and individual learning conditions is treated at a motor, cognitive, and motivational-affective level.

To validate the test score interpretation, we consulted various sources of evidence with regard to the test content, the internal structure of the test, and the interrelationships with conceptually related constructs. The evidence based on the test content is supported by the fact that the item development was systematically derived from the theoretically modeled construct and that the test content was assessed by experts with regard to its fit with the specified construct as well as its representativeness with respect to the theoretical construct. Additionally, the results regarding the response processes from the cognitive lab study indicate that the selected video vignettes represent authentic situations for individual support in inclusive PE and are perceived as such by students. It should be noted that although the assessment of the test content is based on expert judgment, no expert rating was conducted. In order to make this evidence more reliable, further validation steps should be carried out in the future.

The examination of the validity of the response processes shows that it is possible on all eight items to differentiate between individual ability characteristics of noticing with regard to providing individual learning support. Overall, the items were able to provoke situation-specific processes of noticing good fits or misfits between a learning task and individual learning conditions at the motor, cognitive, and motivational-affective levels. Corresponding to the findings of expert-novice studies (e.g., Berliner, 2001; subject-specific: Reuker, 2017a, 2017b), differences in the quality of noticing are documented here.

On the basis of the confirmatory factor analysis with the eight items, the assumption of the unidimensionality of noticing with regard to providing individual learning support can be substantiated. Overall, the scale shows satisfactory psychometric quality and fit with the PC Rasch model. Since the three-dimensional model,

which takes into account the motor-physical, the cognitive, and the motivational-affective levels of individual support, also shows a good fit to the data, the items could also be used for specific investigations of the fit or misfit between a learning task and individual learning conditions at one of the three levels. However, this would require further validation steps.

Moreover, the fit of problems that can be identified in the video vignettes of this instrument are predominantly related to overchallenging situations (in six of eight items). In further studies, it might be beneficial to systematically investigate to what extent there are differences in noticing regarding overchallenging and underchallenging situations.

Furthermore, descriptive statistics show that none of the items had a high solution rate, this applies especially to the items with non-directed focus of attention and solving the items thus tended to be difficult for the sample. In studies investigating correlations between learning conditions and the expression of situated competence facets, different facets of professional knowledge emerge relatively consistently as predictors of noticing relevant classroom situations (e.g., Kersting, 2008; Stürmer et al., 2015). Theoretical declarative knowledge is differentiated from situation-related procedural knowledge. The latter can be understood as the practically usable action knowledge in a concrete instructional situation, which is based on declarative knowledge but is supplemented by personal experiences (Kersting et al., 2010; Seidel & Stürmer, 2014). One reason for the low solution rates could be that comprehensive classifications and explanations require procedural knowledge (Meschede et al., 2015; Seidel & Stürmer, 2014). The validation study presented here was conducted on prospective PE teachers at the beginning of their studies. It may be assumed that this sample had little procedural knowledge, as they lacked suitable learning opportunities in their previous education. Against this background, the low solution rates of the items are within expectations. Further surveys with different groups of experts or with groups with different levels of education should prove useful as an additional validation step with regard to these assumptions. However, the low solution rates with regard to the use of the instrument may also be seen in a positive light, among other things for the evaluation of interventions for promoting noticing with regard to providing individual learning support. Since it is generally assumed that the defined construct can be learned (see also Section 2.2), it is possible due to the low initial scores that the instrument can measure learning progress. However, it has not yet been tested whether the instrument is sufficiently sensitive for this purpose.

The low solution rates on the items might also explain the fact that noticing with regard to providing individual learning support is weakly correlated with learning opportunities in the areas of general inclusion and inclusive PE: None of these learning opportunities builds up the procedural knowledge that is central to the facet of noticing. Rather, these learning opportunities are mostly limited to the teaching of declarative knowledge.

In addition, it can be seen as a limitation that the test items mainly focus on a fit between a learning task and the individual learning conditions of a single student or of different students, which reduces the complexity compared to the real require-

ments for the teacher (many parallel fitting problems that have to be addressed simultaneously in a common teaching situation). However, especially in this first phase of the qualification of prospective PE teachers, this approach also seems to be advantageous, as it allows selective attention to events that are relevant to learning as well as those that hinder it and to interpret them on a theoretical basis without becoming too demanding.

In summary, the findings to date indicate that the test instrument developed here was able to provide a standardized measurement of noticing with regard to providing individual learning support in inclusive PE. On the basis of these pilot studies, the test instrument could be further developed that could be used for the evaluation of a PETE concept based on typical requirement situations of inclusive PE for preservice teachers (Erhorn, Wirszing, & Langer, 2023). As already indicated at various points, however, further validation steps remain necessary for its use in different contexts.

Acknowledgement

This work was supported by the German Federal Ministry of Education and Research under Grant 01NV1712.

Data Availability Statement

Items and the full test scale are available and can be requested from the corresponding author.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Berliner, D.C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Researcher*, *35*, 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Block, M.E., Giese, M., & Ruin, S. (2017). Inklusiver Sportunterricht – eine internationale Standortbestimmung [Inclusive physical education – an international perspective]. *Sonderpädagogische Förderung Heute*, *62*(3), 233–243.

Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education*, *27*(7), 1131–1140. https://doi.org/10.1016/j.tate.2011.04.008

Blömeke, S., Gustafsson, J.-E., & Shavelson, R.J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bruns, J., Strahl, C., & Gasteiger, H. (2020). Situative Beobachtung und Wahrnehmung angehender frühpädagogischer Fachpersonen im Bereich Mathematik – Entwicklung und Validierung eines Testinstruments [Situational observation and perception of pre-service early childhood educators in the field of mathematics – Development and validation of a test instrument]. *Unterrichtswissenschaft*, 1–27. https://doi.org/10.1007/s42010-020-00091-7

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*, 229–238. https://doi.org/10.1023/A:1023254226592

Corno, L. (1995). The principles of adaptive teaching. In A.C. Ornstein (Ed.), *Teaching: Theory into practice* (pp. 98–115). Allyn and Bacon.

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Corno, L., & Snow, R.E. (1986). Adapting teaching to individual differences in learners. In M.C. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 605–629). American Educational Research Association.

Dresel, M., Schober, B., Ziegler, A., Grassinger, R., & Steuer, G. (2013). Affektiv-motivational adaptive und handlungsadaptive Reaktionen auf Fehler im Lernprozess [Affective-motivational adaptive and action adaptive reactions on errors in learning processes]. *Zeitschrift für Pädagogische Psychologie*, *27*(4), 255–271. https://doi.org/10.1024/1010-0652/a000111

Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht [Old wine in new skins? Adaptive teaching: Conceptual reflections]. *Zeitschrift für Erziehungswissenschaft*, *22*(2), 249–277. https://doi.org/10.1007/s11618-018-0840-0

Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion* [Test theory and test construction]. Hogrefe.

Erhorn, J., & Langer, W. (2022). Qualifizierung angehender Sportlehrkräfte für einen inklusiven Sportunterricht [Qualification of prospective PE teachers for inclusive PE]. In D. Katzenbach, M. Urban, D. Lutz, J. Becker, F. Buchhaupt, & A. Strecker (Eds.), *Qualifizierung der pädagogischen Fachkräfte für inklusive Bildung: Vol. 3. Qualifizierung der pädagogischen Fachkräfte für inklusive Bildung – Sekundarstufe // Qualifizierung für Inklusion: Sekundarstufe* (pp. 101–114). Waxmann.

Erhorn, J., Langer, W., & Möller, L. (2020). Förderung und Evaluation von situationsspezifischen Fähigkeiten für einen inklusiven Sportunterricht. Überlegung zu einer zentralen Herausforderung universitärer Sportlehrkräftebildung [Promotion and evaluation of situation-specific skills for inclusive physical education. Reflections on a central challenge of pre-service PE teacher education]. *Qualifizierung für Inklusion*, *2*(1). https://doi.org/10.21248/QfI.33

Erhorn, J., Moeller, L., & Langer, W. (2020). Vorbereitung angehender Sportlehrkräfte auf einen inklusiven Sportunterricht? Eine kritische Bestandsaufnahme hochschuldidaktischer Lehrformate [Academic physical education training for inclusive physical education? A critical review of the state of research]. *German Journal of Exercise and Sport Research*, *50*(4), 487–500. https://doi.org/10.1007/s12662-020-00668-5

Erhorn, J., Wirszing, D., & Langer, W. (2023). Qualification of prospective PE teachers for inclusive PE: Development and evaluation of a PETE concept. *Physical Education and Sport Pedagogy*. Advance online publication. https://doi.org/10.1080/17408989.2023.2194901

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Friesen, M., Kuntze, S., & Vogel, M. (2018). Videos, Texte oder Comics? Die Rolle des Vignettenformats bei der Erhebung fachdidaktischer Analysekompetenz zum Umgang mit Darstellungen im Mathematikunterricht [Videos, texts, or comics? The role of the vignette format for the collection of data about didactic analytical competency when dealing with various media in mathematics class]. In J. Rutsch, M. Rehm, M. Vogel, M. Seidenfuß, & T. Dörfler (Eds.), *Effektive Kompetenzdiagnose in der Lehrerbildung: Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (pp. 153–177). Springer. https://doi.org/10.1007/978-3-658-20121-0_8

Giese, M., & Weigelt, L. (Eds.). (2017). *Edition Schulsport: Band 34: Inklusiver Sport- und Bewegungsunterricht: Theorie und Praxis aus Sicht der Förderschwerpunkte* [Edition school sports: Vol. 34: Inclusive sports and physical education: Theory and practice from the perspective of promotion priorities]. Meyer & Meyer. https://doi.org/10.5771/9783840312304

Hardy, I. (2017). Individuelle Förderung von Lernprozessen – Einführung [Individual promotion of learning processes – Introduction]. In U. Hartmann, M. Hasselhorn, & A. Gold (Eds.), *Entwicklungsverläufe verstehen – Kinder mit Bildungsrisiken wirksam fördern: Forschungsergebnisse des Frankfurter IDeA-Zentrums* (pp. 267–279). Kohlhammer.

Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften: Eine Vertiefungsstudie zur TEDS-Follow-Up-Studie* [Situational diagnostic competence of mathematics teachers: An in-depth study following the TEDS-Follow-Up-Study] [Dissertation]. WorldCat. http://gbv.eblib.com/patron/FullRecord.aspx?p=4518924

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, *68*(5), 845–861. https://doi.org/10.1177/0013164407313369

Kersting, N., Givvin, K.B., Sotelo, F.L., & Stigler, J.W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*(1–2), 172–181. https://doi.org/10.1177/0022487109347875

Knievel, I., Lindmeier, A.M., & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics*, *13*(2), 309–329. https://doi.org/10.1007/s10763-014-9608-z

König, J. (2015). Measuring classroom management expertise (CME) of teachers: A video-based assessment approach and statistical results. *Cogent Education*, *2*(1), Article 991178. https://doi.org/10.1080/2331186X.2014.991178

Kramer, C., König, J., Kaiser, G., Ligtvoet, R., & Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situationsspezifische Fähigkeiten angehender Lehrkräfte [Using videos in teacher preparation: On the effectiveness of video- and transcriptbased classroom management courses on pre-service teachers' general pedagogical knowledge and situation-specific skills]. *Zeitschrift für Erziehungswissenschaft*, *20*(1), 137–164. https://doi.org/10.1007/s11618-017-0732-8

Krammer, K., Hugener, I., Biaggi, S., Frommelt, M., Fürrer Auf der Maur, G., & Stürmer, K. (2016). Videos in der Ausbildung von Lehrkräften: Förderung der professionellen Unterrichtswahrnehmung durch die Analyse von eigenen bzw. fremden Videos [Videos in teacher preparation: Promotion of the professional perception of class by analyzing their own videos or videos of others]. *Unterrichtswissenschaft*, *44*(4), 357–372.

Langer, W., Bruns, J., & Erhorn, J. (2022). Entwicklung und Validierung eines videobasierten Testinstruments zur Erfassung des Noticing mit dem Fokus auf An-

erkennungsprozesse im inklusiven Sportunterricht [Development and validation of a video-based test instrument to measure noticing with regard to processes of recognition in inclusive physical education]. *German Journal of Exercise and Sport Research, 52*(3), 386–397. https://doi.org/10.1007/s12662-022-00807-0

Langer, W., Bruns, J., & Erhorn, J. (in press). Development and validation of a video-based test instrument to measure the situation-specific skills perception and interpretation with regard to providing participation in inclusive physical education. In A. Gegenfurther & R. Stahnke (Eds.), *New perspectives on teacher professional vision*. Routledge.

Lieberman, L.J. (2017). The need for universal design for learning. *Journal of Physical Education, Recreation & Dance*, *88*(3), 5–7. https://doi.org/10.1080/07303084.2016.1271257

Lieberman, L.J., & Houston-Wilson, C. (2018). *Strategies for inclusion: Physical education for everyone* (3rd ed.). Human Kinetics.

Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen [Video-vignette-based standardized collection of data about teachers' cognition]. In U. Riegel & K. Macha (Eds.), *Fachdidaktische Forschungen: Vol. 4. Videobasierte Kompetenzforschung in den Fachdidaktiken* (pp. 45–61). Waxmann.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. https://doi.org/10.1007/BF02296272

Meschede, N., Steffensky, M., Wolters, M., & Möller, K. (2015). Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht: Theoretische Beschreibung und empirische Erfassung [Professional perception of learning support in natural scientific primary education: Theoretical description and empirical acquisition]. *Unterrichtswissenschaft*, *43*(4), 317–335.

Neuber, N., & Pfitzner, M. (Eds.). (2012). *Individuelle Förderung im Sport* [Individual learning support in sports] (2nd ed.). Lit Verlag.

Ní Bhroin, Ó., & King, F. (2020). Teacher education for inclusive education: A framework for developing collaboration for the inclusion of students with support plans. *European Journal of Teacher Education*, *43*(1), 38–63. https://doi.org/10.1080/02619768.2019.1691993

Oser, F. (2013). Kompetenzen der Lehrenden und Auszubildenden: Ein Ressourcenmodell [Competencies of teachers and trainees: A resource model]. In F. Oser, T. Bauder, P. Salzmann, & S. Heinzer (Eds.), *Ohne Kompetenz keine Qualität: Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (pp. 29–65). Julius Klinkhardt.

Pfitzner, M., & Neuber, N. (2012). Individuelle Förderung im Sport – didaktisch-methodische Grundlagen [Individual promotion in sports – didactic-methodological foundations]. In N. Neuber & M. Pfitzner (Eds.), *Individuelle Förderung im Sport* (2nd ed., pp. 75–95). Lit Verlag.

Prengel, A., Riegler, S., & Wannack, E. (2009). „Formative Assessment" als Re-Impuls für pädagogisch-didaktisches Handeln ["Formative assessment" as re-impulse for pedagogical-didactic practice]. In C. Röhner, C. Henrichwark, & M. Hopf (Eds.), *Jahrbuch Grundschulforschung: Vol. 13. Europäisierung der Bildung: Konsequenzen und Herausforderungen für die Grundschulpädagogik* (pp. 253–257). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91721-4_38

Reuker, S. (2017a). 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

Reuker, S. (2017b). The noticing of physical education teachers: A comparison between groups with different expertise. *Physical Education and Sport Pedagogy*, *22*(2), 150–170. https://doi.org/10.1080/17408989.2016.1157574

Reuker, S. (2018). "Ich unterrichte so, wie es die Ereignisse erfordern" – der professionelle Blick von Sportlehrkräften und seine Bedeutung für adaptiven Unterricht

["I teach according to the affordances" – the professional vision of physical education teachers and its importance for adaptive teaching]. *Zeitschrift für Sportpädagogische Forschung*, *6*(2), 31–52. https://doi.org/10.5771/2196-5218-2018-2-31

Reuker, S., & Rischke, A. (2017). Heterogenität im inklusiven Sportunterricht: Situative Deutungsformen und Umgangsformen von Studierenden [Heterogeneity in inclusive physical education: Situational forms of interpretation and handling of students]. *Sportunterricht*, *66*(12), 371–376.

Reuker, S., Rischke, A., Kämpfe, A., Schmitz, B., Teubert, H., Thissen, A., & Wiethäuper, H. (2016). Inklusion im Sportunterricht: Ein Überblick über internationale Forschungsergebnisse aus den Jahren 2005 bis 2014 [Inclusion in physical education classes: An overview of international research results from 2005–2014]. *Sportwissenschaft*, *46*(2), 88–101. https://doi.org/10.1007/s12662-016-0402-7

Rosseel, Y., & Jorgensen, T. D. (2019). *lavaan: Latent variable analysis* (Version 0.6-5) [Computer software].

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensional, and immediacy. *American Educational Research Journal*, *28*(1), 63–88. https://doi.org/10.3102/00028312028001063

Sawalies, J., Veber, M., Rott, D., & Fischer, C. (2013). *Inklusionspädagogik in der ersten Phase der Lehrerbildung. Eine explorative Studie zu Stand und Unterschieden universitärer Lehrangebote für die Regelschullehrämter* [Inclusive education during university studies for becoming teachers. An explorative study of the current situation and differences of university courses for teachers in primary and secondary schools]. *Schulpädagogik heute*, *4*(8).

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „OBSERVER" – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht ["OBSERVER" – Validation of a video-based instrument to record the professional perception of teaching]. *Zeitschrift für Pädagogik*, 296–306.

Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen [How teachers perceive and assess teaching. The recording of pedagogical-psychological competencies with videos]. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik* (Zeitschrift für Erziehungswissenschaft, Special Issue 8, pp. 201–216). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90865-6_12

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

Seidel, T., & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung [Standards and trends in video-based teaching and learning research]. *Zeitschrift für Erziehungswissenschaft, 20*, 1–21. https://doi.org/10.1007/s11618-017-0726-6

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011). Situating the study of teacher noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 3–14). Routledge.

Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, *11*(2), 107–125. https://doi.org/10.1007/s10857-007-9063-7

Steffensky, M., Gold, B., Holodynski, M., & Möller, K. (2015). Professional vision of classroom management and learning support in science classrooms – Does professional vision differ across general and content-specific classroom interactions? *International Journal of Science and Mathematics*, *13*, 351–368. https://doi.org/10.1007/s10763-014-9607-0

Stürmer, K., Könings, K. D., & Seidel, T. (2015). Factors within university-based teacher education relating to preservice teachers' professional vision. *Vocations and Learning*, *8*(1), 35–54. https://doi.org/10.1007/s12186-014-9122-z

United Nations Educational, Scientific and Cultural Organization. (1994). *The Salamanca statement and framework for action on special needs education* (Document No. ED-94/WS/18). https://unesdoc.unesco.org/ark:/48223/pf0000098427

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

van Es, E. A. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 134–151). Routledge.

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Information Technology for Teacher Education*, *10*(4), 571–596.

van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, *24*(2), 244–276. https://doi.org/10.1016/j.tate.2006.11.005

Vygotskij, L. S. (1963). *Thought and language*. MIT Press.

Wagner, L., & Ehlert, A. (2018). Begleitinstrument – diagnostische Kompetenzen [Accompanying instrument – diagnostic competencies]. In A. Borowski, H. Prechtl, & A. Ehlert (Eds.), *Potsdamer Beiträge zur Lehrerbildung und Bildungsforschung: Vol. 1. PSI-Potsdam: Ergebnisbericht zu den Aktivitäten im Rahmen der Qualitätsoffensive Lehrerbildung (2015–2018)* (pp. 73–82). Universitätsverlag Potsdam.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* [Interrater agreement and rater reliability: Methods for determination and improvement of the reliability of assessments by means of systems of categories and rating scales]. Hogrefe.

Wolters, P., Klinge, A., Klupsch-Sahlmann, R., & Sinning, S. (2009). Was ist nach unseren Vorstellungen guter Sportunterricht? [What do we consider good teaching of physical education?] *Sportunterricht*, *58*(3), 67–72.