Analytical Competence of Teachers

Assessing the Construct Validity by Means of Mixed Methods and Drawing Consequences for Teacher Education

Wilfried Plöger, Matthias Krepf, Daniel Scholl, & Andreas Seifert

Abstract

This article focuses on the analytical competence of teachers, which can be defined as the ability to properly perceive and assess lessons in terms of the effectiveness of pupils' learning. However, learning opportunities for fostering this competence in teacher education can only be justified if the components of this construct are theoretically defined and empirically proven. This was the purpose of two studies on which we report here. The studies were combined to a mixed methods design in which the analytical competence of teachers has been investigated. The data were collected using a quantitative method (standardized test; N = 800) as well

Wilfried Plöger is a professor and Matthias Krepf is a research assistant, both at the University of Cologne, Cologne, Germany. Daniel Scholl is a professor at the University of Vechta, Vechta, Germany. Andreas Seifert is a professor at the University of Bielefeld, Bielefeld, Germany. Email addresses are: wilfried.ploeger@uni-koeln.de, matthias.krepf@uni-koeln.de, daniel.scholl@uni-vechta.de, & andreas.seifert@uni-bielefeld.de

as a qualitative method (content analysis of interviews; $N = 18$). The construct of teachers' analytical competence was defined as having two dimensions: a content and a formal dimension. The content dimension is made up of two factors: pedagogical knowledge and content knowledge. The formal dimension of the construct is defined using different levels as factors of complexity with which teachers can process information relevant for teaching. Both studies provided empirical evidence for the validity of the underlying construct. Based on these results, recommendations for teacher education are given.

Introduction

Lesson planning, conducting lessons, and lesson analyses are the core activities of teachers. While the planning phase of the teaching process should anticipate and provide the structure for the lesson to be conducted, its analysis provides the teacher with information on the quality of the planning and the effectiveness of the teaching on students' learning. In this article, we focus on the last aspect, the analysis of lessons as a situation-specific skill (Blömeke, Gustafsson, & Shavelson, 2015), and label the ability to perceive and assess properly one's own and others' lessons in terms of their effectiveness for learning as analytical competence (AC).

Within teacher education, it is considered important to foster teachers' AC, because it is "hard to imagine teachers becoming more effective over time without being able to analyze teaching in terms of its effects on student learning" (Hiebert, Morris, Berk, & Jansen, 2007, p. 48). For this AC to reach a professional level, there is wide consensus that the analysis of lessons is a knowledge-based process and that at least two types of knowledge are necessary prerequisites for an appropriate analysis: pedagogical knowledge (Borko, Koellner, Jacobs, & Seago, 2011; König et al., 2014; van Es & Sherin, 2002) and content-related knowledge (Hiebert et al., 2007; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Star & Strickland, 2008; van den Berg, 2001). Additionally, it is assumed that the AC of teachers depends on the ability to capture the complexity of the whole teaching process. This means that teachers must be able to identify single events as meaningful subactivities in a larger lesson (Berliner, 2001) and to link them to broader didactic units.

Against this background, the construct of teachers' AC could be defined as having two dimensions: a content and a formal dimension. The content dimension is made up of two factors: pedagogical knowledge and content knowledge. The formal dimension of the construct is defined using different levels as factors of complexity with which teachers can process information relevant for teaching. However, little empirical evidence exists whether these components (pedagogical knowledge, content-related knowledge, and the ability to capture the complexity of the whole teaching process) constitute the factorial structure of AC and whether they influence together the degree of teachers' AC.

Research Question

Answering these questions is of theoretical, empirical, and practical relevance, because only if the components of a construct (here AC) are clearly defined and empirically proven can practical consequences for teacher education programs fostering the AC be justified. These three aspects (defining the construct AC by theoretical assumptions, interpreting relevant empirical data in view of these assumptions, and drawing consequences for teacher education) form the structure of this article. First, based on an overview of the literature on research on AC, we theoretically define this construct. Second, we report on our projects in which the AC of in-service and of preservice teachers has been investigated. The interpretation of the data gathered was guided by the research question whether the data matched the theoretical assumptions of the construct.

Examining whether the interpretation of empirical data can be plausibly supported by theoretical arguments is a matter of validity (Kane, 2013; Messick, 1995). However, the question of validity cannot be answered sufficiently in the context of a single empirical study (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Therefore we used a mixed methods approach for implementing and evaluating two educational studies. The quantitative and qualitative data (analytical performance of experts and novices) were collected in the first study using a standardized test and just 2 years later in the second study using a qualitative analysis of in-depth interviews. Finally, based on the results of the two studies, we give recommendations for designing learning opportunities fostering the AC of future teachers.

Theoretical Framework

Research on Analytical Competence of Teachers

Within teacher education, it is considered important that knowledge is a key component of teachers' professional competence (Baumert & Kunter, 2013; Shulman, 1986). Thus the individual form that teachers' AC may take depends on the quality of their knowledge. Regarding the structure of professional knowledge of teachers, current research on the AC of teachers highlights in particular the importance of pedagogical knowledge and content knowledge. Several studies support this assumption (see later), and their empirical data can be interpreted as valid evidence for the importance of these individual knowledge facets for AC.

The importance of pedagogical knowledge for AC was first examined in a qualitative study by van Es and Sherin (2002), whose purpose was to "support teachers in learning to first notice what is significant in a classroom interaction, then interpret that event, and then use those interpretations to inform pedagogical decisions" (p. 575). They made use of video clubs, at which, over the period of around 1 year, a group of teachers regularly met to analyze teaching videos. While at the beginning of the study the ability of the teachers was largely limited to simply noticing teaching situations, by the end they were also able to interpret them by drawing connections between individual events and "broader principles of teaching and learning" (p. 573). These broader principles obviously form an important part of a teacher's pedagogical knowledge.

More recent studies have drawn directly from this approach. For instance, König et al. (2014) examined in a quantitative study the meaning of pedagogical knowledge for an appropriate analysis of teaching situations. The results showed a correlation, $r = .37$, of medium size between pedagogical knowledge and the ability to interpret teaching scenarios but only a small correlation, $r = .17$, between this knowledge and noticing.

The quantitative studies of Seidel and Stürmer (2014) also revealed pedagogical knowledge to be a determining factor for the successful analysis of teaching. From a content point of view, they defined this knowledge using characteristics such as goal orientation, learning assistance, and a positive learning atmosphere. In terms of form, this knowledge enabled analytical performance that could be categorized into three different qualitative levels: description, explanation, and prediction of teaching situations.

While the studies of König et al. (2014) and Seidel and Stürmer (2014) sought to understand the importance of pedagogical knowledge, Kersting and colleagues (Kersting, 2008; Kersting et al., 2012) examined, in the context of mathematics lessons, the role of content knowledge and pedagogical knowledge to reach an adequate analysis. Both forms of knowledge were incorporated within the term *mathematical knowledge for teaching* (MKT). They were able to show that teachers with more comprehensive MKT, compared with teachers who were found to have a lower MKT, were in a far better position to analyze teaching scenarios and to generate alternative teaching strategies (Kersting, 2008). Moreover, the teachers with more comprehensive content knowledge achieved, in terms of its form, a higher level of interpretation, which is determined by the coherence and integration of their comments as well as the recognition of cause–effect relationships connected with the effectiveness of the teaching scenarios on learning outcomes.

In summary, it can be concluded from an examination of the results of this research that the degree of AC in terms of content depends on both pedagogical knowledge and content knowledge. In addition, this knowledge must be organized, in terms of its form, in such a way that not only simple analysis is possible, such as noticing and describing, but also so that more complex information processing, involving interpreting or explaining and predicting, can be achieved.

However, in the studies presented, the influence of individual factors on AC (pedagogical knowledge, content knowledge, and the formal organization of knowledge) was studied separately. Therefore questions relating to how these facets can be integrated into a comprehensive theoretical construct (AC) and how their interactions can be explained in the actual analysis of teaching remain unanswered.

In the following, we give a definition with which these facets are combined to form a coherent model of teachers' AC.

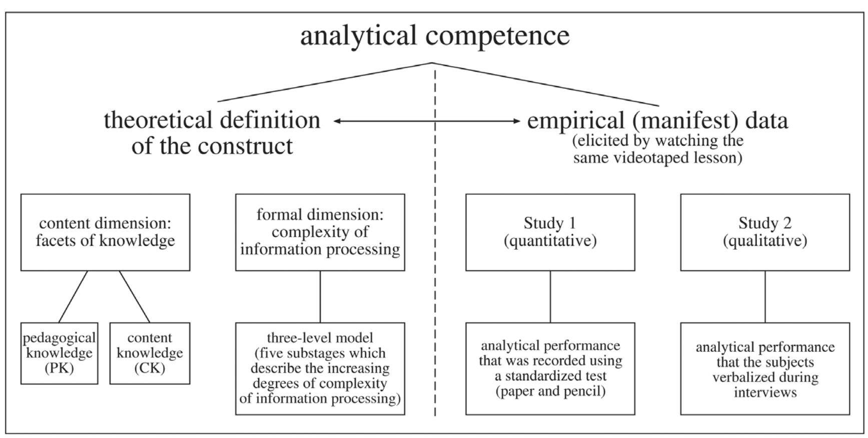
Defining Analytical Competence

The theoretical definition of AC and the chronological sequence of our two studies are presented in Figure 1. The left part of the figure describes the theoretical definition of the underlying construct (including subconstructs, respectively, factors), while the right part shows the quantitative and qualitative data, which were collected in the two consecutive studies.

The individual facets that have been studied in isolation in the aforementioned studies (pedagogical knowledge, content knowledge, and the ability to do more complex information processing) make up the factorial structure of the construct of AC. This construct is defined as having two dimensions: a content and a formal dimension. The content dimension is made up of two factors: pedagogical knowledge and content knowledge. The formal dimension of the construct is defined using three main levels and five substages as factors of complexity with which teachers can process information relevant for teaching.

Pedagogical knowledge. The question of whether teachers are in a position to offer teaching that is effective for learning depends on whether they know and can apply the required general principles of effective teaching. We consider these principles to be an important part of pedagogical knowledge. They not only apply to the planning and implementation of lessons but also must guide the subsequent analysis. Consequently, teachers have a reasonable level of AC if they can recognize

Figure 1
Definition of the construct of analytical competence and chronological sequence of the two studies.



and assess whether the lesson observed has been shaped around these principles and has resulted in the pupils' learning progress.

Proof of the empirical evidence and theoretical plausibility of these principles of effective teaching can be found in many prominent meta-analyses and literature reviews (see, e.g., Brophy, 2000; Hattie, 2009; Kyriakides, Christoforou, & Charalambous, 2013; Muijs & Reynolds, 2011; Seidel & Shavelson, 2007; Walberg & Paik, 2000). These all emphasize the following principles: goal setting/goal orientation; structuring; cognitive activation/motivation; relating cognitive activities to prior knowledge; feedback/evaluation/assessment; adapting/differentiating instruction; using appropriate examples, analogies, and multiple representations; and application/transfer.

Content knowledge. Besides pedagogical knowledge, a high level of AC also requires a high degree of content knowledge. Teachers must know basic concepts and their relationships and master the application of methods that are typical for the individual subject. Moreover, they must be able to teach the content of the subject in such a way that it is understandable for pupils. The content knowledge necessary in each case depends on the specific topic of the lesson under consideration. In the case of the two studies we carried out (standardized test, interviews), the respondents need physical knowledge from the field of optics, because they are to analyze a videoed lesson on the subject of light refraction.

Complexity of information processing. We describe the formal quality of knowledge using various levels of AC. To this end, we consider the analysis of teaching as a case of information processing and refer to the *model of hierarchical complexity* that was developed by Commons and others (Commons, Goodheart, Dawson, & Draney, 2008) for all the areas in which information must be processed in an organized form. This is also the case for the analysis of teaching.

Organized information processing is necessary if the difficulty of processing increases with an increasing number of content elements. Such challenges cannot be overcome using simple, almost additive chains of elements but rather by using nonarbitrary links, in which the degree of complexity processed is expressed. The gradual manifestation of the necessary complexity can be described in stages. Higher stages arise from the linking of elements below, so that across the stages, the complexity increases.

This content-independent model of hierarchical complexity has great potential for the area of competence research. Nevertheless, it is necessary to adapt it in specific ways so that it can describe the construct AC. This is done here through the formulation of a five-stage model (Figure 2), whose focus is on the increasing level of complexity of the information to be processed.

The model is divided into three main levels (Roman numerals) and five substages (Arabic numerals). Stages 1 and 2 limit the measurement of AC to individual teaching events (visible actions and/or situations). People whose analytical performance corresponds to these two stages are therefore only able to perceive and describe individual events within the lesson and to determine the degree to which they exist.

From Stage 3 upward, the processing of information is complex enough to be able to relate individual actions to the effects that they cause (Stage 3) or to connect individual actions to broader didactic units and to provide alternatives for such units (Stage 4). In relation to creating alternatives, a greater effectiveness for learning can then be predicted in respect of the expected effects.

Stage 5 encompasses the observation and evaluation of the entire teaching process. People who achieve this level are able to assess the importance of didactic units against the background of the whole lesson and specify justifiable alternatives to such units.

Methods

Context of the Studies

The two studies presented here were arranged sequentially and collected data of the respondents through different methods (see the right side of Figure 1). In the first study, the AC of the respondents was measured via a quantitative method using a standardized test (paper and pencil). In the second study, the respondents analyzed the lesson being observed during focused one-on-one interviews. These verbal data were then categorized and counted.

Both studies were expected to result in empirical data that would offer information about the factorial structure of the construct according to the definition given

Figure 2
Stages of analytical performance showing the necessary complexity level of information processing.

	Level III
Stage 5	The teacher – can evaluate the importance of didactic units with regard to the *whole* teaching process and provide justified alternatives for such units
	Level II
Stage 4	– can relate individual behaviors or situations to *broader didactic units* and provide alternatives for such units
Stage 3	– can identify or predict with a high level of certainty the effects or outcomes that *individual* behaviors or situations can or will bring about
	Level I
Stage 2	– can determine the *degree of the manifestations* of these categories
Stage 1	– can classify *individual* teaching events (behaviors or situations) into the relevant categories

earlier. The samples were chosen so that people with very high (experts) and very low analytical performance (novices) were represented. In the sample for Study 2, people who had already participated in the first study and had proven their expertise in the analysis of teaching by means of a standardized test could be included.
The starting point for both studies was the same video stimulus watched by the respondents. To find an appropriate video, seven scholars working in the field of teacher education examined nearly 20 available videos by several criteria. The duration of the video presentation should not exceed a total of 20 min, because the standardized test as well as the interview took up already a considerable amount of time (60 min). However, the video should not represent a collage of individual unconnected scenes but a coherent lesson process to cover all five levels of our competence model (see Figure 2). Moreover, the video should display typical weaknesses regarding the effectiveness for students' learning to provide opportunities for the participants to activate their content and pedagogical knowledge.

Finally, we found a video that met these criteria. This video showed a physics lesson in the field of optics, in which the law of refraction (Snell's law) was being taught. The lesson was 45 min long in total, but the video producers had reduced it down to a 15-min clip by cutting out two phases (carrying out the experiment in group work and performing calculation exercises in partner work). Instead, short displayed text passages faded in between the previous and the following sections described the goals and the progress of these two phases, so that the viewer could imagine what happened in the meantime. Therefore the scholars were in complete agreement that the impression of the entire lesson's process had not been lost. Additionally, they concurred that this lesson represented typical weaknesses of physics lessons as described in the study of Seidel and Prenzel (2006), so that its effectiveness on learning was limited (no clear goal orientation, little cognitive activation, unstructured discussion, inappropriate examples used for demonstration, etc.).

Study 1 (Quantitative)

Sample. Study 1 included prospective and practicing teachers ($N = 800$) who were already teaching or would go on to teach in gymnasium schools (high school track). In Germany, there are three different types of secondary schools: *Gymnasium*, *Realschule*, and *Hauptschule*. Students at *Gymnasium* graduate after Grade 12 with the Abitur, the highest school-leaving certificate, which is also required to enter university-based teacher education. Students at *Realschule* and *Hauptschule* graduate after Grade 10, and then they usually undergo a 3-year apprenticeship program combined with instruction in a part-time vocational school.

The professional development of teachers who work in gymnasium schools can be divided into four phases: They first take two subjects at university in different combinations (e.g., mathematics and history or English and biology) as well as courses in educational studies. Following their studies, they undergo an 18-month practical phase as a trainee teacher (in-school program). Then they teach their subjects independently as in-service teachers. The school administration then selects particularly qualified people from among these teachers to be responsible for the training of trainee teachers as teacher training personnel.

To reproduce these four stages of professional development, the total sample comprised four subpopulations with 200 each of students, trainee teachers, teachers, and teacher training personnel. We assumed before the start of the study that the students would be the novices and that the teacher training personnel would be the experts.

Test instrument. The test used in Study 1 was a quantitative achievement test (paper and pencil) and comprised 23 closed and open-ended items, the responses to which represented the analysis of the lesson. The items were designed such that solving them required predominantly pedagogical knowledge or predominantly content knowledge. Furthermore, the items illustrated the three main levels of the competence model (see Figure 2) and thus represented different degrees of complexity with which information could be processed. For illustrative purposes, we have quoted three items for the three main levels of the competence model whose solutions required predominantly pedagogical knowledge (Items 13 and 6) *or* predominantly content knowledge (Item 22):

> Item 13, Level I
> Does the teacher have an overview of whether his pupils are able to follow the individual steps of the lesson, or not?
>
> Notes: To solve this item, pedagogical knowledge about the importance of feedback is necessary. The crucial function of feedback is that both pupils and teachers receive information relevant for learning that enables them to assess the progress of the learning process undergone so far. In the physics lesson being analyzed, the teacher barely makes use of the opportunities for feedback as, for instance, he does not pose any questions to the pupils or does not call upon them to separate the important aspects from the unimportant ones or to summarize the partial knowledge acquired so far.
>
> Item 22, Level II
> At the end of the lesson, the teacher shows a Fresnell lens as a practical application. Is the physical function of this lens comprehensible for the pupils, or does it remain incomprehensible for them?
>
> Notes: To solve this item, content knowledge about the function of a Fresnell lens is necessary. For instance, it is necessary to know that in producing such a lens, a significant part of the lens material can be saved, because the light rays falling perpendicular to the surface do not undergo any refraction; they only do so after exiting the lens.
>
> Item 6, Level III
> In terms of its process, teaching can be viewed as a structured sequence of scenarios. Within this there are some scenarios that stand out as *key scenarios*.

Such key scenarios determine the progression of teaching events in a positive or negative way. What do you believe are the key scenarios in this lesson (provide a maximum of two key scenarios)?

Notes: The solution of this item requires pedagogical knowledge concerning cognitive activation and the use of appropriate examples. An example of a key scenario is the introduction phase of the lesson. It is of central importance for the entire subsequent teaching process because the pupils are not clear about the actual problem. They know from the previous lesson that light spreads out in a straight line. But in the case of light refraction, the ray changes its direction when it travels through different optical materials. This fact could have been used by the teacher to produce cognitive dissonance among the pupils. However, this cognitive dissonance is not achieved because none of the concrete examples demonstrated at the beginning showed a structure isomorphism with the law of light refraction. As a consequence, later on in the lesson, the pupils are unable to form meaningful hypotheses or construct a suitable experiment.

Data analysis. The analysis of the data was undertaken in three steps, as follows.

1. First, the test was scaled with Conquest software (Wu, Adams, & Wilson, 1997). Following item response theory (IRT), a partial credit model was calculated. The item parameters with associated fit values, the person parameters, and the discrimination indices were determined using this procedure.

2. The person parameters calculated using the IRT scaling were then used to carry out group-specific comparisons with analyses of variance (in particular, between novices and experts).

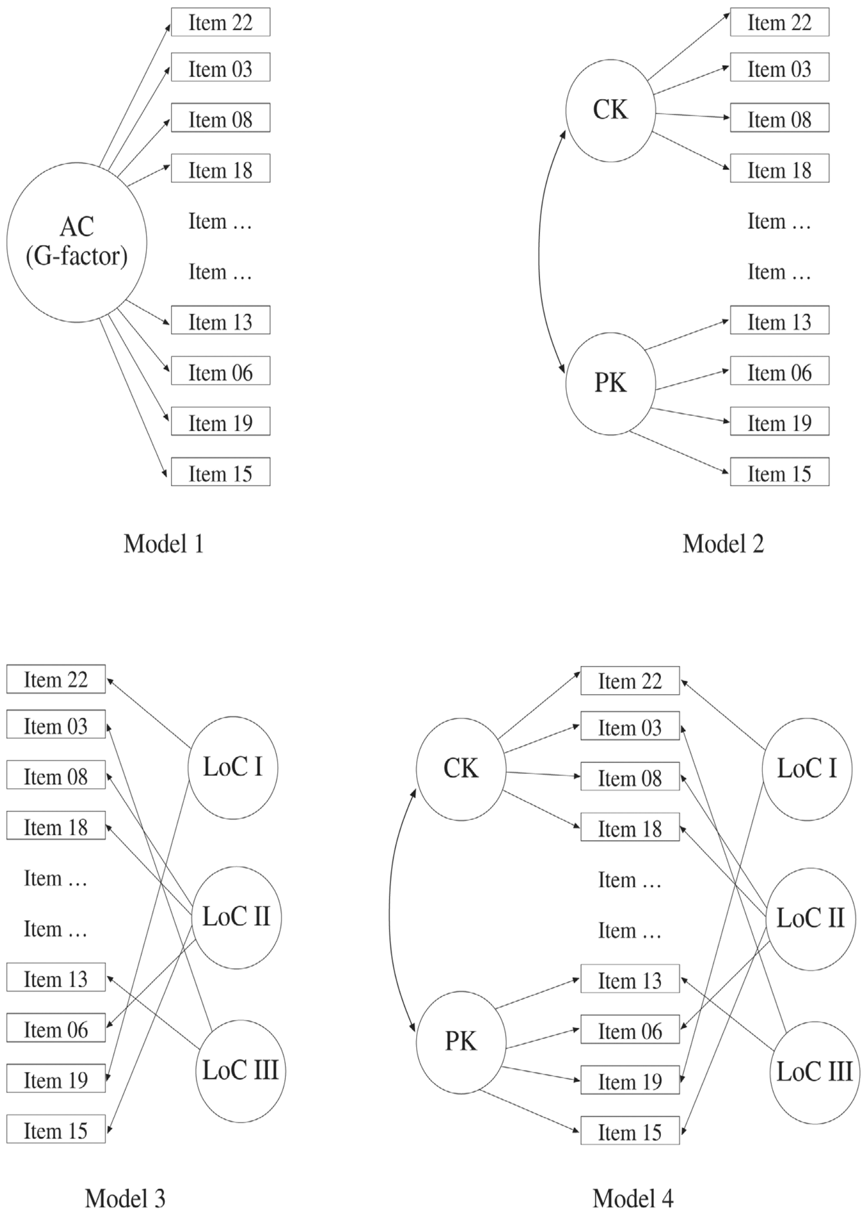
3. To be able to assess the factorial structure of the construct of AC, confirmatory factor analyses were carried out. In these analyses, the two dimensions on the left in Figure 1 and their associated factors form the latent variables to which the test items as manifest measurement indicators were assigned. Depending on the combination of dimensions and factors (see Figure 3), four models arose from a heuristic perspective: Model 1, AC as a G-factor, in which the variance of all manifest variables is explained by a general factor; Model 2, where the items each load on one factor of the content dimension, so on the pedagogical or content knowledge; Model 3, where the items each load on one of the three factors (= main level) of the formal dimension; and Model 4, in which, unlike Models 2 and 3, double loadings on the content and formal dimension are allowed. Each item represents both the content dimension (pedagogical knowledge or content knowledge) and the formal dimension (degree of complexity of the information to be processed). In Model 4, the aspects studied separately in the aforementioned studies on AC are integrated into an overall construct.

Study 2 (Qualitative)

Sample. To ensure a certain comparability of the samples, Study 2 also should involve people who were already teaching or those who will go on to teach in

Figure 3

Modeling analytical competence by combining dimensions and associated factors in different ways. CK = content knowledge. LoC = level of complexity, depth of understanding. PK = Pedagogical knowledge.



gymnasium schools and who differ significantly in the level of their AC. As Study 1 showed, students achieved low analytical performance, while teachers and teacher training personnel showed very high performance. These performance differences in the sample of Study 2 were created by including a group of novices (9 students) and a group of experts (9 teachers and teacher training personnel).

The in-service teachers and teacher training personnel were considered experts regarding three criteria. They should have extensive teaching experience (at least 15 years), their qualification has been recognized by the school administration by promoting them from common teachers to teacher training personnel, and they should have achieved particularly high analytical performance in the standardized test for Study 1 ("top scorers"). Contact with these people was made through the institutions at which they work. They then surrendered their anonymity and voluntarily took part in an interview.

Unfortunately, it was not possible to use novices (students) from the first study for participation in the second study. This is because, for data protection reasons, in the university context, the identity of people who have previously taken part in a study anonymously cannot be identified retrospectively. Therefore, for the second study, we found students who were enrolled at our own university, who participated voluntarily.

Videotaped lesson as stimulus for conducting interviews. The stimulus used to conduct one-on-one interviews with the participants was the aforementioned recording that showed a physics lesson on optics in which Snell's law of refraction was being studied. In the subsequent interview, subjects were asked to comment on the lesson they had observed. To encourage the comments to be as open as possible, the definition of conversation's structure in terms of both time and content was kept to a minimum. The video was shown in five sections. After each section, the interview was initiated with an open question (e.g., "What did you observe in this section of the lesson?"). If the conversation came to a halt, further encouragement for discussion was given (e.g., "How could the teacher's behavior that you just described affect the pupils?"). At the end of the interview, the participants were asked to once again summarize what they believed the most important aspects of the lesson observed were.

Data analysis. The analysis of the data was undertaken in two steps. First, the recorded interviews were transcribed in full and analyzed using the qualitative content analysis method (Mayring, 2014). This method allowed extensive text material to be reduced down to essential statements and structures. This works on the assumption that there are suitable categories within which the relevant text passages can be coded. As there was a need for valid evidence for the factorial structure of the construct of AC (see Figure 1), the categories had to reflect the content dimension (pedagogical and content knowledge) and the formal dimension of the construct (depth of understanding).

Overall, 21 categories were used: (a) 10 categories identifying pedagogical knowledge (e.g., goal orientation, structuring, cognitive activation, feedback, using appropriate examples), (b) six categories for content knowledge (e.g., formulation of physics concepts, conducting experiments as a key content-related method, further content-related working methods), and (c) five categories for the depth of understanding, to record the various levels of complexity of information processing (see Figure 2).

Each of the 21 categories was defined as well as being made concrete by using a prototypical text passage (anchor example). Such anchor examples show analogously how a text passage should be formulated to be included in the appropriate category. These definitions and anchor examples together formed a category system (Mayring, 2014) and specified the guidelines for performing the content analysis.

The performance of the content analysis, which was carried out in part by the second author of this article and two students, required proper and precise handling of the individual categories. Therefore it was necessary to prepare the two students for this task with an intensive training course. After these two people had acquired sufficient certainty in the use of the 21 categories, the actual coding process began. Over a certain period, each of the three people coded all 18 interviews separately. In the meantime, meetings were held to determine where there were similarities or differences in the codings. Ultimately, only the statements on which the three coders reached a consensus were accepted.

The category system used was an important prerequisite not only for the process of coding itself but also for the assessment of the stability and the reproducibility of the codings, which needed to show an acceptable level of intercoder reliability and intracoder reliability. After the coding process had finished, two other people were involved in the process to determine the intercoder reliability. For this process to be carried out using the time as efficiently as possible, these people did not code all the interviews, instead taking on just a selection of the text passages. This selection covered 10% of the codings that had previously been identified per consensus which was reached between the original three people. In accordance with Fleiss and Cohen (1973), a kappa between .60 and .75 was set as a sufficient level for the intercoder reliability. Five weeks later, these two people once again coded the text passages that they alone had worked on to see if they reproduced the same codings as the first time. This intracoder reliability was also allocated a kappa between .60 and .75.

In the second stage, the codings registered were quantitatively processed. Each individual statement coded in the content analysis represented—in analogy to the right response to an open-ended question on a standardized test—a "strike" and could be considered as 1 point. Consequently, the total number of codings recorded for one person represented the extent of his or her own personal knowledge activation. And vice versa, the quantitative differences in the number of codings recorded for different people also reflected quantitative differences in the extent of their knowledge activation. The content analysis therefore provided interval-scaled data that could be processed further quantitatively. In accordance with this, the category system formed a scale that included 21 variables (categories) with three part scales (pedagogical knowledge, content knowledge, depth of understanding).

On the basis of this interval-scaled data, we calculated the discrimination indices for the individual variables and the reliability of the three part scales to assess whether the codings determined by the content analysis could be interpreted as sufficiently precise and coherent "measurements" across all 21 variables. Furthermore, we used the Mann–Whitney test (because of the small sample and the fact that the data were not normally distributed) to check whether the experts and novices differed significantly in their analytical performance.

Quantitative Results

Rasch Analysis of the Test

The EAP/PV reliability (comparable with Cronbach's alpha) of the test used was at .86 and deemed to be good; however, a higher variance of the overall construct (.52) would have been desirable. The weighted mean squares (MNSQs) of the items were between 0.83 and 1.11. Only one item was slightly outside the acceptable range between 0.8 and 1.2, with a MNSQ of 1.22. The discrimination indices were between .28 and .68 and were therefore also acceptable.

The Rasch model is defined, among other things, by the restriction of the *specific objectivity*, which states that the difficulty of a specific item (i.e., item parameter) should be independent (i.e., objective) of the studied populations. The question of whether this ideal requirement of the Rasch model was more or less satisfied was answered using the correlations of the item parameters (see Bond & Fox, 2007) calculated for the four (sub-)populations (with 200 each of students, trainee teachers, teachers, and teacher training personnel). There were consistently high correlations (min. 0.77, max. 0.99), which proved that the requirement of specific objectivity had been met.

Group Comparisons

In the result of the analyses of variance, the main effect of group membership on the level of the test scores achieved turned out to be significant, $F(3, 796) = 84.55$, $p < .001$, $\eta^2 = .24$. In subsequent analyses, it emerged that students (novices) differed significantly from the other three groups. The most significant difference was in the comparison with the group of teacher training personnel (experts). In the context of the well-known group comparison (Cronbach & Mehl, 1955; DeVellis, 1991; Rubin & Babbie, 2015), this result can be interpreted as evidence for the validity of the construct of AC.

Factor Analysis

The fit indices displayed in Table 1 showed no sufficient model fit for the first three models. This is because, in the case of a good fit, the χ^2 value normalized to the degrees of freedom (= χ^2/df) should be between 0 and 2, the RMSEA between 0 and 0.05, and the CFI between 0.97 and 1.00 (Byrne, 2010). By contrast, for the first two values, Model 4 showed a very good fit, and in terms of the CFI value, the fit (.96) was also acceptable.

Thus the results for Model 4 showed that AC could be interpreted as a two-dimensional construct: Only when the factors of the content dimension (pedagogical knowledge and content knowledge) and the factors of the formal dimension (three levels of complexity of information processing) were combined into one model were the fit values sufficiently adequate. The empirical data could therefore be interpreted as clear evidence for the factorial validity of the construct of AC (see the left part of Figure 1).

Qualitative Results

Results of the Content Analysis

There were high values for the repeatability and stability of the codings. Indeed, the intercoder reliability achieved a kappa of .73, and the intracoder reliability was .77.

Table 2 presents the number of codings that were recorded. On the left part of the table are the factors for the content and formal dimensions (content knowledge, pedagogical knowledge, and depth of understanding). The cells in the following columns represent the number of codings that each group (experts vs. novices) achieved for the relevant factor of the content or formal dimension. Overall, 588 text passages representing the respondents' independent activation of knowledge were coded. The considerable number of codings showed that the 21 categories used were a valid way of illustrating the dimensions and factors of the construct of AC.

To make it clear what content statements stood behind these bare figures, we quote specific statements from different interviews that were coded with the corresponding categories and represented the participants' individual (a) pedagogical

Table 1
Goodness-of-Fit Statistics for the Four Models

Model	x^2	df	p	x^2/df	CFI	RMSEA
1	657.83	152	0.000	4.33	0.81	0.07
2	657.10	151	0.000	4.35	0.81	0.07
3	652.15	149	0.000	4.20	0.82	0.06
4	239.30	129	0.000	1.86	0.96	0.03

knowledge and (b) content knowledge and their (c) ability to do more complex information processing.

Statements that represent the activation of pedagogical knowledge (cognitive activation, goal orientation, feedback, structuring). These included the following:

> ***Cognitive activation.*** He is not at all able to focus the pupils' attention. He could also have filled in the image on the board together with the pupils. Therefore it passes the pupils by. It was also noticeable that the ability to concentrate had gone. Some of the pupils looked at the issue as if bored.
>
> ***Goal orientation.*** He seemed to me to be very vague in his whole approach to teaching. Personally, I had the impression that the teacher himself did not have a structured approach in his head with which to direct the pupils where he wanted them to go. He did actually verbalize an objective, but the transparency of how the objective would be reached was not easily distinguishable for me.
>
> ***Feedback.*** [The teacher] himself does not know at all what he has taught the pupils, or what they have understood. They could have looked at what they should do. You could then allow one pupil to summarize. He also does not check at any point how things are going. So this also means that he doesn't see the different learning and performance abilities of his pupils.
>
> ***Structuring.*** "So this is a crucial point, that the phases of the lesson are actually completely isolated and that it is therefore of course also difficult for the pupils to follow the thread of the lesson."

Statements that represent the activation of content knowledge (conducting experiments as a key content-related method, further content-related working methods). These included the following.

> ***Conducting experiments as a key content-related method.*** The scientific path of knowledge could have been considered by the pupils [themselves]: how can we tackle this phenomenon, we want to fathom it out as physicists, we want to discover a law that governs it. How do we do that? He does not take them with him on this scientific path of discovery, he just makes them reproduce it. [And then the] next opportunity he fails to use: How will we design our experiment? The students try

Table 2
Number of Codings for the Content and Formal Dimensions

	Experts	Novices	Total
Content dimension			
Pedagogical knowledge	180	97	277
Content knowledge	97	44	141
Formal dimension			
Depth of understanding	134	36	170
Total	411	177	588

to introduce an idea, which is also essentially shelved very quickly. Therefore, he did not consider encouraging autonomy in scientific thinking at all in this sequence, I see that as a very extreme shortcoming in this lesson.

Further content-related working methods (graphical illustration of measurement data; consideration of measurement errors). He missed many opportunities here. One example of such an opportunity: At school we naturally do not undertake a large error analysis. Nevertheless, we think about measurement errors. There is actually no better way than using the variations in the pupils' experiments as a measure of the precision. So he missed that here. And the second, where I believe a huge opportunity was missed: Even as experts we are not able to discern much from a table. Therefore the first step is actually always to depict the whole thing in a graph. You have to do this!

Statements that represent the depth of understanding. The following statements relate to four points at which the video presentation was interrupted and the teacher was asked to give his opinion. Immediately after the first sequence, he tried to predict how the rest of the lesson would go, even though he did not know at this stage. In the following statements, the teacher constantly referred back to the previous sequences. This shows that he understood the course of the entire lesson as a unit. This level of understanding corresponds to Stage 5 of the model (understanding the *entire* teaching process) presented in Figure 2.

Following the first sequence. "It is an attempt to develop the issue of refraction as a problem jointly with the pupils. However, it is possible to see already that he cannot stick to this principle."

Following the second sequence. Basically it continues in this way. Normally at this point he should ask what results they had got, taking the opportunity to allow the pupils to join in the conversation, or to exchange their values. Once again he does not take what the pupils could contribute seriously.

Following the third sequence. "Now he tries again to get the pupils on board with him, but it is actually too late."

Following the last sequence. That was once again a very interesting sequence, because it confirmed what was indicated from the beginning. He is able, on the one hand, to use his humorous nature to repeatedly engage with the pupils on a personal level, but at the same time he does here precisely what he has already been doing the whole time.

Reliability of the Codes/Homogeneity of the Scales and of the Construct

Table 3 shows the calculated reliability values. The reliability of the 16 variables (categories) of the content dimension was high at .82. The value for the scale of the formal dimension (depth of understanding) was .62. That was still an acceptable value, however, because this scale was only made up of five variables.

The correlations between the incidences of coded content for the factors of the content and formal dimension (see Table 4) were consistently high. This supported the fact that the three part scales (variables for pedagogical knowledge, content knowledge, and depth of understanding) added to a homogeneous overall construct

Group Comparison

The implementation of the Mann–Whitney test showed that the experts differed significantly from the novices in all three part scales (pedagogical knowledge, $U = 12{,}5$, $z = -2.48$, $p = 0.011$; content knowledge, $U = 12{,}5$, $z = -2.48$, $p = 0.011$; depth of understanding, $U = 0{,}00$, $z = -3.60$, $p < 0.001$). As proposed by Fritz, Morris, and Richler (2012), the effect size r of the Mann–Whitney test can be calculated from the z-values in relation to the size of the sample ($N = 18$) by the formula $r = z / \sqrt{N}$, whereby r is comparable with Pearsons's correlation coefficient. The calculated effect sizes were $r = .58$, .58, and .85. The differences between experts and novices were therefore also of practical importance, because according to Cohen (1988), values ≥ .5 represent a high effect.

Summary and Discussion

Both studies reported here were focused on the research question of whether the data generated in different ways could be coherently interpreted with the theoretical assumptions of the underlying construct (AC). Regarding this research question, it can be summarized that the analysis of the quantitative and qualitative data led

Table 3
Reliability Values of the Categories for the Three Part Scales

	Cronbach's alpha	No. of categories
Content dimension		
Pedagogical knowledge	.74	10
Content knowledge	.66	6
Pedagogical + content knowledge	.82	16
Formal dimension		
Depth of understanding	.62	5

Table 4
Correlations Between the Incidences of Coded Content of the Three Part Scales

	Pedagogical knowledge	Content knowledge
Content knowledge	.68**	
Depth of understanding	.56*	.66**

*$p < .05$. **$p < .001$.

to positive and consistent answers.

The confirmatory factor analyses used for the evaluation of the quantitative data confirmed the two-dimensional construct. Only when the factors of the content dimension (pedagogical knowledge and content knowledge) and the factors of the formal dimension (three main levels of complexity of information processing) were combined into one model were the fit values acceptable. The evaluation of the qualitative data led to a large amount of coded content and proved that the 21 categories meaningfully represented the two dimensions and the associated factors of the construct. Both the acceptable reliability values for the coded content and the high correlations between the incidences of coded content for the three part scales supported the homogeneity of the construct. In conclusion, therefore, we note that the factorial structure of the construct of AC identified in the first study using quantitative data was validly replicated in the second study using qualitative data.

Another important result was based on the composition of the samples of both studies. They were organized such that in every sample was a group of experts and novices who were expected to differ significantly in the level of their AC. If the data collected later correspond with these expectations, this well-known group comparison could be interpreted as evidence for the validity of the underlying construct (Cronbach & Mehl, 1955). This was the case for the evaluation of both the quantitative and qualitative data. As regards the calculated test scores, the analyses of variance carried out in Study 1 revealed significant differences between the performances of the novices and the experts. This result was also confirmed by the Mann–Whitney test carried out in the second study.

Regarding the differences between the experts and novices in this second study, it could be argued that the experts had an advantage, because they had seen the video a second time after just 2 years, while the novices had seen it for the first time. However, we assume that possible learning effects were rather small, because the experts (9 top scorers out of 800 people) had already shown in Study 1 their extraordinarily high AC, which seems to be hardly increasable by viewing the video a second time. Moreover, expert teachers can be characterized by *subitizing*, that is, the rapid, accurate, and confident judgment of teaching events at first glance (Berliner, 2001), so that a second viewing may not significantly improve their analytical achievements.

Beyond the complementarity of the reported results (confirming the structure of the construct and the differences between experts and novices), two further findings emerged that can be attributed to the specific strengths of the two approaches, through which they are able to examine certain aspects of the assessment of the construct validity better than the other approach. The use of the quantitative instrument provided precise information on the current achievement of the respondents. Owing to the relatively large sample ($N = 800$), cautious conclusions on the generalizability of the test results could be drawn in connection with the proof of the specific objectivity of the item scale. We assume that the quantitative instrument used is a fundamentally appropriate way of reliably diagnosing the AC of teachers (at least the performance of those who teach in gymnasium schools).

In contrast to this strength of the test instrument, its limitation regarding the assumed knowledge activation must be seen. The items used implicitly represented the pedagogical or content knowledge that we as researchers found relevant for solving the items (see the preceding item examples). However, the question of whether the knowledge independently activated by the respondents in the test was congruent with the researchers' imputed knowledge could not be answered with the quantitative procedure but rather by the qualitative approach. The statements of the participants could be coded into many categories that represented the two dimensions and the associated factors of the construct. This can be seen, for instance, in the fact that during the analysis of the lesson, pedagogical knowledge for the assessment of goal orientation, cognitive activation, feedback, or structuring was activated (see the preceding quotations). The same also applies for the activation of content knowledge. Ultimately, the statements of the respondents were able to illustrate the extent to which they could capture the complexity of the entire teaching process. Taken together, the visualization of knowledge facets and the documentation of understanding processes differing in complexity reflect the strength of the qualitative approach.

Recommendations for Teacher Education

On the basis of the results of our two studies and the confirmed structure of AC, some recommendations for designing teacher education programs and special learning opportunities fostering the AC of future teachers can be made. First, preservice teachers have to acquire a sufficient level of content-related knowledge, especially subject matter knowledge. Having this kind of knowledge at one's disposal does not automatically mean that one will be able to use it for teaching and analyzing lessons (Ball, 2000; Lampert, 2010). On the contrary, it must be deconstructed so that the crucial knowledge elements are accessible and visible for the pupils. Only then is it also effective for the analysis of lessons and helpful in identifying lesson goals, assessing whether the goals are being achieved, constructing hypotheses about cause–effect relationships between teaching and students' learning, and using analysis to propose improvements in teaching (Hiebert et al., 2007; Kersting et al., 2012). Particularly suitable seem to be learning opportunities, in which the acquisition of both these analytical skills and content knowledge is combined, as Roth et al.'s (2011) study with science teachers has shown.

Second, preservice teachers must be given sufficient opportunity to acquire the necessary pedagogical knowledge on principles of effective teaching (e.g., goal orientation, cognitive activation, feedback, structuring, clarity). To avoid these principles remaining isolated knowledge elements, it must be shown that they together form a coherent unit. For this purpose, these principles may be centered on the theory of meaningful learning (Fiorella & Mayer, 2015) so that preservice teachers understand why the teaching behavior must correspond to these principles. To illustrate this, we give two examples. First is relating cognitive activities to prior knowledge: To be able to build a new cognitive structure, new knowledge elements always need to be connected to prior knowledge elements. Second is using appropriate examples: Succinct examples make it easier to pupils to distinguish important from irrelevant elements.

To ensure the acquisition of the necessary knowledge of principles of effective teaching, education at university should fulfill at least three criteria. First, similar to the video studies reported earlier, it should be based on concrete contexts, provide the principles of effective teaching, and decontextualize the knowledge of these principles through a variety of applications (Korthagen, 2010). This procedure may reduce the risk of pedagogical knowledge remaining idle and being lost when entering the profession. Second, as well as teaching in schools, education at university should follow the principles of effective teaching and should make the thought processes of the prospective teachers the central reference point. This should enable teacher educators to demonstrate that these principles are not empty statements but feasible practices. Finally, the results of such courses should be evaluated both quantitatively and qualitatively to ensure that the overarching objective (i.e., the acquisition and application of academic knowledge about effective teaching as a necessary instrument for lesson analysis) is achieved in reality.

We followed these criteria in a recent study (Plöger, Scholl, & Seifert, 2018), which took as its basis the implementation of a curriculum (treatment) through which preservice teachers had the opportunity to acquire theoretical knowledge about the principles of effective teaching and to apply this knowledge to the analysis of videotaped lessons of others. The evaluation was carried out using a Solomon four-group design ($N = 316$). From the pre- to the posttest, the participants in the experimental group (acquisition of knowledge of principles of effective teaching), compared with the participants in the control group, showed a significant increase in their AC with a medium effect size.

An appropriate analysis of lessons requires not only content-related and pedagogical knowledge but also the ability to capture the complexity of teaching, as the comparison between novice and expert teachers in our two studies have shown. This result is in line with other studies. Novice teachers tend to notice classroom situations step by step and to perceive lessons as a chronological but disconnected sequence of events (Star & Strickland, 2008). As a result, they cannot recognize and assess the functions of lesson elements for the whole teaching process (Berliner, 2001). In contrast, experienced teachers do not consider elements of lessons in isolation but think systematically about their interrelations (Rosaen, Carlisle, Mihocko, Melnick, & Johnson, 2013). They identify single events as meaningful subactivities within a larger lesson. Therefore they are able to link various single situations and actions to broader didactic units (phases) and to evaluate the effectiveness of such units against the background of the whole teaching process.

These significant differences point to specific development tasks to improve the analytical skills of preservice teachers within teacher education. To pick up preservice teachers where they are, it is necessary to support them "in learning to first notice what is significant in a classroom interaction, then interpret that event, and then use those interpretations to inform pedagogical decisions" (van Es & Sherin, 2002, p. 575). For this purpose, priority should be given to short videos in combination with predetermined observation tasks focusing on a limited number of principles of effective teaching to avoid observers' cognitive overload. As the certainty has increased to identify and assess single events, more complex situations should be chosen for making connections visible between single teaching events. In this case, videos of longer duration are suitable, because they capture more complex situations, similar to real-life teaching scenarios.

Acknowledgments

This work was supported by the German Research Foundation (grant number PL 272/3–1].)

References

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

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education, 51*, 241–247. https://doi.org/10.1177/0022487100051003013

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Kraus, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-5149-5_2

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

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

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM Mathematics Education, 43*(1), 175–187. https://doi.org/10.1007/s11858-010-0302-5

Brophy, J. (2000). *Teaching* (Educational Practices Series No. 1). Retrieved from http://www.ibe.unesco.org/

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applica-*

tions, and programming (2nd ed.). New York, NY: Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Commons, M. L., Goodheart, E. A., Dawson, T. L., & Draney, K. (2008). Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. *Journal of Applied Measurement, 9*, 182–199.

Cronbach, L. J., & Mehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302. https://doi.org/10.1037/h0040957

DeVellis, R. F. (1991). *Scale development: Theory and application*. Newbury Park, CA: Sage.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613–619. https://doi.org/10.1177/001316447303300309

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology, 141*, 2–18. https://doi.org/10.1037/a0024338

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon, UK: Routledge. https://doi.org/10.4324/9780203887332

Hiebert, J., Morris, A., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education, 58*, 47–61. https://doi.org/10.1177/0022487106295726

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73. https://doi.org/10.1111/jedm.12000

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

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*, 568–589. https://doi.org/10.3102/0002831212437853

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38*, 76–88. https://doi.org/10.1016/j.tate.2013.11.004

Korthagen, F. A. J. (2010). Situated learning theory and the pedagogy of teacher education: Towards an integrative view of teacher behavior and teacher learning. *Teaching and Teacher Education, 26*, 98–106. https://doi.org/10.1016/j.tate.2009.05.001

Kyriakides, L., Christoforou, C., & Charalambous, Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*, 21–34. https://doi.org/10.1177/0022487109347321

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American*

Psychologist, 50, 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Muijs, D., & Reynolds, D. (2011). *Effective teaching: Evidence and practice* (3rd ed.). London, UK: Sage.

Plöger, W., Scholl, D., & Seifert, A. (2018). Bridging the gap between theory and practice: The effective use of videos to assist the acquisition and application of pedagogical knowledge in preservice teacher education. *Studies in Educational Evaluation, 58*, 197–204. https://doi.org/10.1016/j.stueduc.2017.12.009

Rosaen, C. R., Carlisle, J. F., Mihocko, E., Melnick, A., & Johnson, J. (2013). Teachers learning from analysis of other teachers' reading lessons. *Teaching and Teacher Education, 35*, 170–184. https://doi.org/10.1016/j.tate.2013.06.007

Roth, K. J., Garnier, H., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. (2011). Videobased lesson analysis. *Journal of Research in Science Teaching, 48*, 117–148. https://doi.org/10.1002/tea.20408

Rubin, A., & Babbie, E. R. (2015). *Essential research methods for social work* (4th ed.). Belmont, CA: Brooks/Cole Cengage Learning.

Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction, 16*, 228–240. https://doi.org/10.1016/j.learninstruc.2006.03.002

Seidel, T., & Shavelson, R. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454–499. https://doi.org/10.3102/0034654307310317

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal, 51*, 739–771. https://doi.org/10.3102/0002831214531321

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

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

van den Berg, E. (2001). An exploration of the use of multimedia cases as a reflective tool in teacher education. *Research in Science Education, 31*, 245–265. https://doi.org/10.1023/A:1013193111324

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*, 571–596.

Walberg, H. J., & Paik, S. J. (2000). *Effective educational practices* (Educational Practices Series No. 3). Retrieved from http://www.ibe.unesco.org/

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software* [Computer program]. Camberwell, Australia: Australian Council for Educational Research.