American Educational Research Journal
August 2014, Vol. 51, No. 4, pp. 739–771
DOI: 10.3102/0002831214531321

Modeling and Measuring the Structure of Professional Vision in Preservice Teachers

Tina Seidel
Kathleen Stürmer
Technische Universität München

Professional vision has been identified as an important element of teacher expertise that can be developed in teacher education. It describes the use of knowledge to notice and interpret significant features of classroom situations. Three aspects of professional vision have been described by qualitative research: describe, explain, and predict classroom situations. We refer to these aspects in order to model professional vision. We developed a video-based instrument to empirically test the model. The results show that our measure to assess aspects of professional vision differentiates between description, explanation, and prediction. The study provides insight into the structure of professional vision, allowing us to conceptualize it theoretically and discuss the targeted use for teaching and formative assessment of preservice teachers.

KEYWORDS: teacher education, teacher expertise, professional vision, video, item response theory, competence assessment

The way teachers design and create learning opportunities in their classrooms strongly influences student learning (Darling-Hammond & Bransford, 2005; Goldstein & Hersen, 2000; Seidel & Shavelson, 2007). Thus, defining and measuring competencies that teachers require for creating those learning opportunities are of particular importance in teacher

---

TINA SEIDEL is a professor of teaching and learning research at the School of Education of the Technische Universität München, Marsstr. 20-22, 80335 München, Germany; e-mail: *tina.seidel@tum.de*. Her research interests include video-based analysis of classrooms, teaching adaptivity and aptitude-treatment-interaction, and development of multimedia-based tools in teacher research.

KATHLEEN STÜRMER is a research scientist at the School of Education of the Technische Universität München, Marsstr. 20-22, 80335 München, Germany; e-mail: *kathleen.stuermer@tum.de*. Her research interests include empirical research on professional development of teachers, diagnostic of competencies, and development of multimedia-based tools in teacher research.

education (Brouwer, 2010; Cochran-Smith, 2003; Koster, Brekelmans, Korthagen, & Wubbels, 2005).

So far, however, only limited empirical research exists with regard to the structure and the development of teacher competencies over time. Research in teacher education is still quite a "young" field (Grossman & McDonald, 2008). Compared to teacher education, intensive efforts have been made in other professions to achieve advancement in the measurement of professional competencies (Boshuizen, Schmidt, Custers, & van de Wiel, 1995; Shavelson, 1991). For a long time most instruments in teacher education research used teachers' subjective judgments of their abilities or rather distal indicators such as degrees, courses taken, or certificates. There has been critical discussion about these available measures (cf. Voss, Kunter, & Baumert, 2011; Wayne & Youngs, 2003). In this context, modeling the structure of teacher competencies and testing models empirically is regarded as helpful in order to advance the field (Koeppen, Hartig, Klieme, & Leutner, 2008).

As a theoretical background, teacher education researchers have drawn on Shulman's (1987) conceptualization of teacher knowledge and theories of teacher learning (Borko, 2004), taking into account the situational and contextualized nature of teacher knowledge. In line with Shulman's ideas, some approaches have focused on providing empirical evidence for the proposed structure of teacher knowledge (Baumert et al., 2010; Hill, Rowan, & Ball, 2005; Voss et al., 2011). Furthermore, researchers have taken a stance in focusing on the situational and contextualized nature of teacher learning, describing processes of learning that might lead to the proposed structure of teacher knowledge (Hammerness, Darling-Hammond, & Shulman, 2002; Putnam & Borko, 2000).

In both contexts, the use of video has become a prominent tool for studying teacher learning and the activating of teacher knowledge (Brophy, 2004; Goldman, Pea, Barron, & Denny, 2007; Kersting, 2008). Many approaches in video-based teacher research have their foundation in expertise research and have identified different qualities of teacher knowledge (Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987; Putnam & Borko, 2000). Teachers' reasoning about video-based examples of classroom practice serves as an indicator for the quality of teacher knowledge. High reasoning abilities indicate differentiated and integrated knowledge with a flexible application to various teaching situations. Low reasoning abilities on the other hand indicate fragmented and rather sparse knowledge structures without the ability to use this knowledge flexibly.

Many of these studies are embedded in the context of in-service teacher professional development. Their focus is on describing individual changes in teachers' knowledge or the development of teacher groups reasoning jointly about video—for example, in "video clubs" (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Sherin & van Es, 2009; van Es, 2009). Although these studies measure teachers' reasoning in a contextualized manner, standardized conditions for assessing individual teacher knowledge are rarely achieved.

These kinds of measures, however, would be helpful for teaching and formative assessment purposes. Qualitative approaches as used in research so far require quite some time and effort. This has the disadvantage that results can be used as feedback only after a significant amount of time. Quantitative instruments as the one presented and proposed in this article allow more efficient data analysis. Such a measure can provide a first indicator regarding, for example, the current state of teacher competencies. These indicators can be used promptly for feedback on teaching and formative assessment. In the long run, the combination of qualitative and quantitative measures would be an ideal objective. So far, however, quantitative measures are underdeveloped compared to the rich body of knowledge stemming from qualitative approaches.

A promising approach developed by Kersting (2008) has responded to this challenge by using standardized video clips as "item prompts." These prompts are embedded in open questions that tap into teachers' individual interpretations of classroom situations. Kersting's findings show that the standardized use of video clips to measure teachers' ability to analyze classroom situations represents a valid approach to assess their knowledge.

Whereas significant advances in studying competencies of in-service teachers have been made, *initial* processes of knowledge acquisition and teacher competence development as provided by university-based teacher education have not been studied intensively (Brouwer, 2010; Koster et al., 2005). Integrating theory and practice in order to support the initial acquisition of teacher knowledge is particularly important in this context (Cochran-Smith & Zeichner, 2005).

A promising indicator for "integrated" teacher knowledge is seen in the concept of professional vision (Goodwin, 1994). For teachers, professional vision is the ability to notice and interpret relevant features of classroom situations (van Es & Sherin, 2002). Preservice teachers at the beginning of their university-based education are often not able to direct their attention to relevant elements of classroom instruction in order to select events and situations that have been shown to influence student learning (Star & Strickland, 2008). In fact, many preservice teachers struggle when they begin their first teaching post, finding themselves unable to deal with the complexity of the classroom environment and to apply what they have learned at university to the context of the classroom (Stokking, Leenders, de Jong, & van Tartwijk, 2003). Therefore, one of the key aims of preservice teacher education is to systematically foster the acquisition of integrated knowledge as indicated by professional vision.

So far, teachers' professional vision has been studied by qualitative approaches and descriptions of teachers' approaches toward noticing and reasoning. These findings have provided a valid basis for describing the quality of teacher knowledge and teacher learning. Yet, in order to move the field forward, professional vision should also be investigated by using diverse methodological approaches, for example, by testing models (in which previous findings are summarized) empirically and using quantitative methods. Therefore, we decided to expand the research in accordance with this desideratum in our approach. We use the descriptive knowledge gained from qualitative research to model the structure of professional vision and to test it empirically, the aim being to (a) advance the scientific understanding of professional vision by modeling and empirically testing certain characteristics of the construct in the context of university-based teacher education and (b) develop and provide a contextualized, yet standardized instrument that—in the long run—can be used as a formative assessment instrument in university-based teacher education.

In this article, we report the most significant findings of this larger research project targeting the question whether our research can be regarded as a valid and reliable approach to assessing some characteristics of professional vision. These findings serve as an important basis for additional questions of the research project, for example, whether participation in university-based courses on teaching and learning leads to positive developments of preservice teachers. We outline this larger research project in order to provide a framework for integrating the findings of the present study and discuss them in more detail at the end of the article. In the next section the conceptual framework for modeling the structure of professional vision with a focus on university-based teacher education is summarized.

Modeling the Structure of Professional Vision

Noticing: Prompting Teacher Knowledge by the Selection of Classroom Events

Given the situated and contextualized nature of teacher knowledge, one aim in developing a measure to assess professional vision is to use authentic video sequences of classroom situations as "prompts" to elicit teacher knowledge (Kersting, 2008). Even in short sequences of classroom teaching, a myriad of teaching and learning acts occur. Some are of particular importance for student learning, others are not. In this vein, the situations and events teachers direct their attention to while observing a classroom sequence serve as a first indicator for the activation of teacher knowledge. In line with the concept of professional vision, we refer to noticing as a relevant component of professional vision (Sherin, Jacobs, & Randolph, 2011). Noticing describes whether teachers pay attention to events that are of importance for teaching and learning in classrooms, for example, influencing student learning in a positive or negative way.

When it comes to defining such relevant events, knowledge about principles of teaching and learning (Grossman & McDonald, 2008) as an aspect of generic pedagogical knowledge (Shulman, 1987) provides an important frame. Generic pedagogical knowledge is required to create learning environments across a wide variety of subjects (Voss et al., 2011) in a domain-general manner (Blomberg, Stürmer, & Seidel, 2011). It represents a basic component of initial university-based teacher education (Hammerness et al., 2002; Voss et al., 2011), and the majority of all preservice teachers are required to acquire this kind of knowledge. Teaching effectiveness research is based on knowledge about teaching and learning as an element of generic pedagogical knowledge. In this research a number of teaching and learning (TL) components have been repeatedly shown as relevant for student learning (Fraser, Walberg, Welch, & Hattie, 1987; Goldstein & Hersen, 2000; Hattie, 2009; Seidel & Shavelson, 2007). Seidel and Shavelson (2007), for example, used a cognitive process-oriented teaching and learning model to summarize findings of the past decade. The model includes the components of goal setting and orientation, execution of learning activities, evaluation of learning processes, teacher guidance and support, and the social environment in which learning takes place. Each component shows differential effects on student learning, including cognitive and motivational-affective aspects.

Since a first attempt to create a standardized yet contextualized measure as proposed in our research would be overburdened if this full model is applied, we decided to focus on three components (see Figure 1): *goal clarity and orientation, teacher support and guidance*, and *learning climate*. The three are selected since they represent a balance of important teaching and learning components. Learning climate served as an indicator for the motivational-affective classroom context, goal clarity and orientation for the successful preparation of learning, and teacher support as a guiding process involved in the execution of learning activities. The three components were repeatedly identified as relevant for student learning. The component of goal clarity and orientation (clarifying teaching and learning goals, structuring the lesson), for example, is particularly relevant with regard to cognitive and motivational aspects of student learning, since students should activate their knowledge and be motivated to learn. Teacher support and guidance positively affects student learning, particularly with regard to motivational-affective aspects. Teacher questions, as well as their reactions to student responses in the form of feedback, are core elements of research in this area. The learning climate in a classroom is of particular relevance for student learning since it provides an important motivational and affective background in which learning takes place. Two aspects that have been addressed in this strand of research are teacher humor as well as teachers taking the needs of students seriously.

Teaching effectiveness research provides an important basis for selecting classroom situations that show effects on student learning. However, this research has limitations since the fine-grained processes that mediate between teaching and learning are not investigated. In order to provide evidence-based reasoning of noticed classroom events, models and knowledge about these processes are important. In this context, teaching

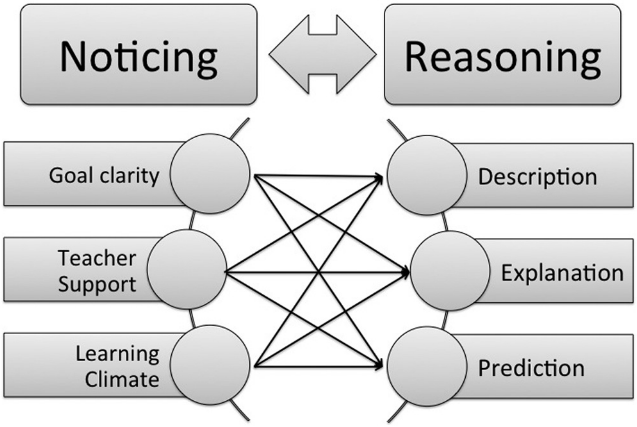

Figure 1. **Noticing and reasoning as two components of professional vision.**

effectiveness researchers refer to self-determination theory (SDT) in order to model the processes involved in the creation of learning environments by teachers and the effective use by learners. SDT proposes three basic conditions that a learning environment needs to address in order to make learning processes likely: the experience of competence, autonomy, and social relatedness (Deci & Ryan, 2002). A substantive body of research has shown that the perception of these conditions in a learning environment is positively related to intrinsic motivation and human development. With regard to the three selected teaching and learning components as derived from teaching effectiveness research, it has been shown that goal clarity and orientation is an important condition for students to experience competence, autonomy, and social relatedness (Kunter, Baumert, & Koller, 2007; Seidel, Rimmele, & Prenzel, 2005a), with positive effects on student motivation and knowledge development over time. In addition, teacher support and guidance in classroom discourse is positively related to the three conditions with positive effects on intrinsic learning motivation and interest development (Lipowsky et al., 2009; Seidel, Rimmele, & Prenzel, 2003). Furthermore, a positive learning climate positively affects the perception of the three conditions, again with positive effects on student learning (Buff, Reusser, Rakoczy, & Pauli, 2011).

In summary, classroom situations that are relevant for teaching and learning can be selected on the basis of empirical evidence as provided by teaching effectiveness research that is based on generic pedagogical knowledge. This selection process provides the framework for the activation of teacher knowledge, indicated by what preservice teachers notice while observing situations. However, it is also obvious that the complexity of classroom teaching and learning goes beyond what can be measured with a single instrument. Thus, decisions about selections have to be made, and the interpretations being drawn from the use of the instrument have to be restricted to the context in which the measure is placed (Kane, 1994). In our context we decided to focus on the three components, goal clarity and orientation, teacher support, and learning climate, stemming from the model of Seidel and Shavelson (2007). Thereby, we wanted to assure that the instrument can be expanded later to the full process-oriented teaching and learning model if this first attempt to develop a quantitative measure is empirically supported. The three components represent a substantial part of the teaching and learning model. To expand the instrument to the full model, the two further components of executing learning activities and evaluation would have to be included.

Reasoning as an Indicator of "Integrated" Teacher Knowledge

In the previous section, we referred to noticing as the first of two important subcomponents of professional vision (van Es & Sherin, 2002). In this section, we summarize research on the second subcomponent: teachers' reasoning about classroom sequences (see Figure 1). The ability to take a reasoned approach to events noticed in the classroom provides insight into the quality of teachers' mental representations of knowledge and the application of those representations in the classroom context (Borko, 2004; Borko et al., 2008). When it comes to conceptualizing teachers' reasoning, research often distinguishes three qualitatively different aspects (Berliner, 2001; Borko & Livingston, 1989; Sherin & van Es, 2009; van Es, 2009): (a) description, (b) explanation, and (c) prediction. The three aspects were derived from observing teachers while they reflected on videotaped classroom situations and are based on qualitative analysis. For example, Sherin and van Es (2002) recorded preservice teachers discussing videotaped classroom situations over the course of their video club interventions and coded their statements according to those categories. These aspects are conceptualized as distinct but highly interrelated (van Es & Sherin, 2002).

Description refers to the ability to clearly differentiate the relevant aspects of a noticed teaching and learning component (i.e., goal clarity) without making any further judgments. It is an important aspect of reasoning to use professional knowledge to describe the situation before explaining the situations and predicting the possible consequence. Taking the example of goal clarity, a person observing the first minutes of a lesson might state that the teacher refers to what the students should learn, how the lesson is structured, and how the contents relate to what the students have learned previously.

Explanation refers to the ability to use what one knows to reason about a situation. This means linking classroom events to professional knowledge and classifying situations according to the components of teaching involved. With regard to goal clarity, one would expect a person to link the observation to professional terms and concepts, for example, stating that the teacher is clarifying teaching and learning goals and activating students' pre-knowledge.

Prediction refers to the ability to predict the consequences of observed events in terms of student learning. It draws on broader knowledge about teaching and learning as well as its application to classroom practice. With regard to goal clarity, a person might use knowledge about effects of goal clarity on student learning in order to make a prediction about possible consequences. If a teacher, for example, misses clarifying learning goals, a consequence might be that the students are less likely to direct their learning toward the goals with negative consequences for motivation and knowledge acquisition.

To date, research has shown that preservice teachers enrolled in university education programs are capable of describing classroom situations. In contrast, their ability to adequately explain and predict the consequences and outcomes of those situations lags behind that of experienced in-service teachers (Oser, Heinzer, & Salzmann, 2010; Seidel & Prenzel, 2007). In terms of the underlying knowledge structures, these findings indicate that preservice teachers lack the elaborated and integrated knowledge structures that would allow them to link observed situations with knowledge about teaching and learning (Putnam & Borko, 2000). Thus, it can be hypothesized that tasks requiring explanation and prediction are more difficult than tasks requiring description of classroom situations. It might even be that predicting is even more difficult than explaining. Since experienced in-service teachers and school inspectors have been shown to demonstrate all these abilities (Seidel & Prenzel, 2007), it seems likely that the knowledge structures required can be developed over time. However, little empirical research has systematically explored whether preservice teachers develop professional vision as a single ability or if they make progress in the different dimensions of professional vision at different rates.

To our knowledge, no research has been conducted so far to empirically test assumptions about the interrelation of the three aspects of professional vision (describe, explain, predict). For example, professional vision might be regarded as one-dimensional so that the three aspects cannot clearly be separated; it might also be that the three aspects have to be seen as distinctive but highly interrelated. Taking into account the higher-order knowledge application processes involved (Resnick, 1987) and the results of previous studies (e.g., van Es & Sherin, 2008), it also seems possible that explaining and predicting are so closely related that they can be treated as one aspect (as integration). However, none of the research so far indicates that professional vision should be regarded as a second-order factor to occur independent of the three aspects of description, explanation, and prediction. All three aspects are described as being to some extent interrelated. Overall, the structure of reasoning as a central subcomponent of professional vision is not elaborated to a full extent. It remains an open question whether teachers' reasoning depends on the scope of knowledge about teaching and learning that has to be integrated and transferred to an observed situation.

Yet, knowledge about these processes would advance the field, especially when it comes to designing learning environments in university-based teacher education. If the three reasoning aspects are highly interrelated and represent distinctive dimensions of increasing difficulty, teacher educators could draw on this knowledge in order to structure and sequence courses on teaching and learning (Brouwer, 2010; Koster et al., 2005). Therefore, in the present study we empirically tested which operationalization best fits the structure of reasoning when modeled with data on preservice teachers and to what extent these abilities are interrelated.

Research Questions

In this study three research questions are addressed:

Research Question 1: Are the selected video clips discernible examples of the three TL components? Answers to the first research question are of importance with regard to the validity of the selected video clips in the sense that they represent discernible teaching and learning components.

Research Question 2: Does a model operationalizing professional vision into three dimensions (description, explanation, and prediction) fit the data generated by our measure better than a one- or two-dimensional model? Answers to the second research question will show to what extent our measure is functioning as designed: to elicit the three dimensions of description, explanation, and prediction.

Research Question 3: To what extent is the measurement of preservice teachers' professional vision stable over time, showing that no developments or changes occur without further knowledge-based interventions? The answer to this question indicates to what extent the instrument could be used formatively in the context of university-based teacher education.

Method

In the following, we first summarize the development of the video-based instrument. Second, we describe samples and research designs of three consecutive studies in which we tested our research questions. Third, we summarize the instruments used in the three consecutive studies.

Development of the "Observer" Instrument as a Measure for Professional Vision

The development of the instrument called "Observer: Video-Based Tool to Diagnose Teachers' Professional Vision" (Seidel, Blomberg, & Stürmer, 2010b) involved three major parts: (a) the selection of video clips that represent situations relevant for goal clarity, teacher support, and learning climate as prompts to elicit teacher knowledge; (b) the development of ratings connected to the selection of video clips tapping into teachers' reasoning in the three aspects of description, explanation, and prediction. Both parts were then (c) integrated into an online data base.

Noticing: Selection of Video Clips

The selection of video clips for eliciting teacher knowledge about teaching and learning was based on three criteria.

First, given the situated and contextualized nature of teacher knowledge, the selected video clips should be perceived by participants as authentic examples of classroom practice. To achieve this, we decided to use classroom sequences of the educational system that our preservice teachers would encounter, in our case German-speaking classrooms. Since the instrument was supposed to capture generic pedagogical aspects of professional vision, different subjects to which the generic knowledge should be transferred were represented. Given this first criterion, available video recordings of German-speaking instruction in various subjects (e.g., Reusser, 2005–2009) were screened.

Second, the video clips should serve as prompts to activate teacher knowledge. Thus, on the one hand, the selection of video clips should be perceived as stimulating and activating. On the other hand, the video clips should not involve too much complexity, which would lead to increased cognitive load. The objective was to have a balance in the activating nature of the video clip while at the same time not overwhelming observers with high cognitive load.

Third, we focused on teaching effectiveness research in order to identify sequences that are of particular relevance for student learning, either in the way that a positive example is represented (positive example) or in the way that a teacher lacks to address a relevant component (negative/ambiguous example). Theoretical conceptualizations and video coding schemes for the three components of goal clarity, teacher support, and learning climate were used to identify video sequences (Seidel, Prenzel, & Kobarg, 2005), resulting in a pre-selection of 86 video clips. In this process the research team learned that it was hardly possible to identify video clips only representing one of the three components. To account for this, the decision was made to identify video clips representing two of the three components.

In addition, each component was specified with regard to two content aspects (goal clarity: the teacher clarifies the learning goals and the structure of the lesson; teacher support: the teacher asks open questions and gives supportive feedback; learning climate: the teacher uses humor in his or her instruction and takes the needs of the students seriously). Three experts of the research team independently assigned video clips to these specified components. Then, the clips and their assignment were discussed and validated by the experts.

Finally, 12 video clips (each 2 to 4 minutes) covering different subjects (2 × physics, 2 × mathematic, 4 × history, 1 × French, and 1 × English as foreign language) were selected that met the three main criteria. All video clips featured German-speaking Grade 8 and Grade 9 classrooms with students aged between 14 and 16 years.

The authenticity as well as the cognitive load of the video clips were investigated in a pilot study with $N = 40$ voluntary participating preservice teachers (Seidel et al., 2010a). With regard to the use of the instrument in the context of university-based teacher education, we aimed to reach a range of preservice teachers. Thus, the academic experiences of the preservice teachers in the pilot study differed widely (semester $M = 6.69$, $SD = 2.60$). We asked each pilot study participant to think aloud while watching the 12 video clips and to evaluate them with regard to stimulation and mental effort in a short questionnaire after each clip. The think-aloud protocols were analyzed qualitatively. The results showed that overall the video clips were perceived as authentic. Acknowledging that the video clips should be balanced, we compared them with regard to their stimulation ($\alpha = .61$; example item: How did you feel while observing the instruction in the excerpt: It was stimulating) and mental effort ($\alpha = .97$; example item: How did you feel while observing the instruction in the excerpt: It was hard work). Analyses of variance showed that the video clips did not differ in the perception of the participants, stimulation: $F(11, 264) = .87$; $p = .57$; $\eta^2 = .04$; mental effort: $F(11, 264) = 1.10$; $p = .38$; $\eta^2 = .04$. In addition, no significant differences between video clips were found regarding the represented subject, stimulation: $F(1, 274) = .06$; $p = .81$; $\eta^2 < .01$; mental effort: $F(1, 273) = .54$; $p = .59$; $\eta^2 < .01$; and assigned TL component, mental effort: stimulation: $F(3, 272) = 1.29$; $p = .28$; $\eta^2 = .01$; $F(3, 272) = 1.50$; $p = .22$; $\eta^2 = .02$.

Reasoning: Construction of Rating Items

Based on our conceptual framework (Figure 1), rating items were constructed for the combination of TL component and reasoning aspect (Figure 2).

For each of the video clips 36 rating items were developed. A 4-point Likert scale ranging from 1 (*disagree*) to 4 (*agree*) was used. Participants were asked to what extent they agree with the items after having observed a video clip. Table 1 exemplarily shows the rating items for a video clip that represented goal clarity and learning climate.

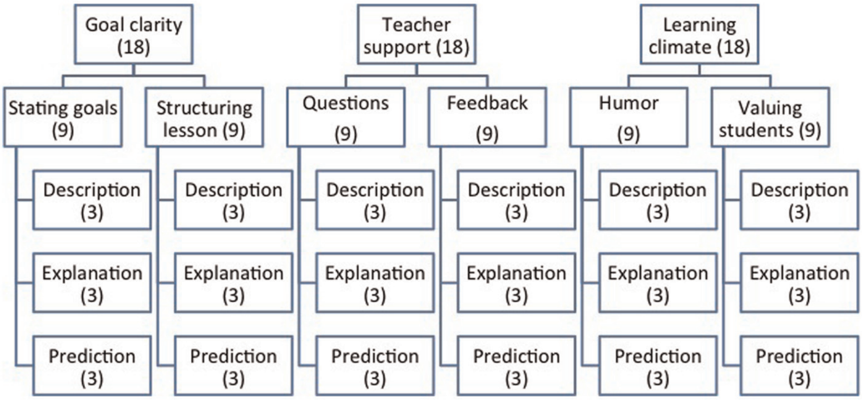

Figure 2. **Rubric for the construction of ratings (number of items).**

In developing rating items, we sought to ensure local independence of responses, a key assumption of item response theory (Embretson & Reise, 2000) on which our data analysis was based. Questions measuring *description* targeted the specific and differentiated observation of the three TL components. Questions tapping into *explanation* focused on the link between an observed event and knowledge about the corresponding TL component, specifically with regard to how a teaching component addresses students' individual perception of the supportiveness (e.g., autonomy, competence) of a classroom situation. Questions assessing *prediction* focused on the possible consequences of an observed situation on student learning, including consequences for learning motivation, cognitive processing, and affect.

Given the fact that the participants' responses are based on subjective ratings of observed TL components and that teaching effectiveness research does not provide right or wrong answers with regard to the quality of video clips, it is necessary to establish a suitable norm or frame of reference. The use of criterion-referenced norms is regarded as suitable, especially in cases in which the variance in a target population is assumed to be low (Goldstein & Hersen, 2000). This is the case for the context of university-based teacher education in which preservice teachers have limited possibilities to acquire professional vision compared to expert teachers. Criterion-referenced norms use content-related criteria such as expert norms for comparison (Oser et al., 2010). This approach is based on the assumption that experts are characterized by having acquired well-structured and integrated knowledge that they use while accomplishing a professional task such as reasoning about an observed video clip (Ericsson, Krampe, & Tesch-Romer, 1993; Kalyuga, 2007).

Table 1
Examples of Rating Items Tapping Into Reasoning

Aspects of Reasoning	Rating Items
Goal clarity: Clarifying the learning goals	
Description	
In the excerpt that you saw . . .	the teacher clarifies what the students are supposed to learn.
	the teacher states the topic of the lesson.
	the teacher places the topic within a broader context.
Explanation	
In the excerpt that you saw . . .	the students have the opportunity to activate their prior knowledge of the topic.
	the students have the opportunity to see the significance of the topic to them personally.
	the students have the opportunity to adopt the teacher's objectives as their own learning goals.
Prediction	
Based on what you saw . . .	the students will be able to align their learning process to the learning objective.
	the students will be able to get acquainted with the topic.
	the students will be able to prepare for what's coming.
Learning climate: Teacher takes students' needs seriously	
Description	
In the excerpt that you saw . . .	the teacher is respectful of the students.
	the teacher shows that he values the students.
	the teacher asks questions/sets tasks that are appropriate for the students' level of development.
Explanation	
In the excerpt that you saw . . .	the students have the opportunity to feel that their teacher takes them seriously.
	the students have the opportunity to contribute substantively in discussions with their teacher.
	the students have the opportunity to develop their own ideas on the material covered.
Prediction	
Based on what you saw . . .	the teaching style will motivate the students.
	the students will be mentally engaged.
	the students will be able to feel at ease in the lesson.

Therefore, we used expert judgments as a criterion norm to measure professional vision in our instrument. To establish this norm, three expert researchers—each with 100 to 400 hours of experience in observing classroom situations according to the TL components under investigation—independently rated all developed rating items in connection with the selected video clips. Cohen's kappa (*k*) was calculated to determine the consistency of the expert ratings; a mean Cohen's *k* of .79 across the three raters indicated a satisfactory level of consistency (Seidel et al., 2010a). In cases where the experts disagreed, agreement was reached by consensus validation.

Based on this criterion-referenced norm, the participants' responses were compared to expert ratings. Depending on the strictness of using the criterion-referenced norm, two ways of calculating "agreement with experts" were set up: (a) a strict measure of 0 (miss expert rating) and 1 (hit expert rating) and (b) a close approximation of 0 (miss expert rating), 1 (correct direction on the scale), and 2 (hit expert rating). Since previous research did not provide information with regard to what strategy fits best in the context of teacher education, we tested our assumptions about the structure of professional vision using both reference norms. Results are reported in the next section.

Integration of Video Clips and Rating Items: Observer-Instrument

The video clips and rating items were integrated into an online tool called "Observer: Video-Based Tool to Diagnose Teachers' Professional Vision" (Seidel et al., 2010b). The online tool is presented as a series of HTML pages and starts with general instructions and short introductions of the three TL components (Figure 3, parts a and b). Video clips are then presented, followed by the rating items targeting TL components and reasoning aspects (describe, explain, predict) (Figure 3, parts c and d). Brief contextual information about the class is provided before each video clip is presented. Participants have the opportunity to watch the clips a second time before responding to the rating items. The final version of the Observer instrument includes a set of six video clips as prompts and accompanying rating items tapping into teacher reasoning. In this form, the completion time of the instrument is about 90 minutes. In sum, the instrument consists of 216 items.

Participants in Three Consecutive Studies

In order to answer our research questions, three consecutive studies were run. The first study focused on answering the first research question, addressing the extent to which the selected video clips are valid examples of TL components and serve the function of eliciting preservice teachers' knowledge. The second study was conducted in order to test model assumptions about the structure of professional vision (second research question). The third study targeted the retest reliability of the instrument over time and whether learning effects in completing the instrument occur.

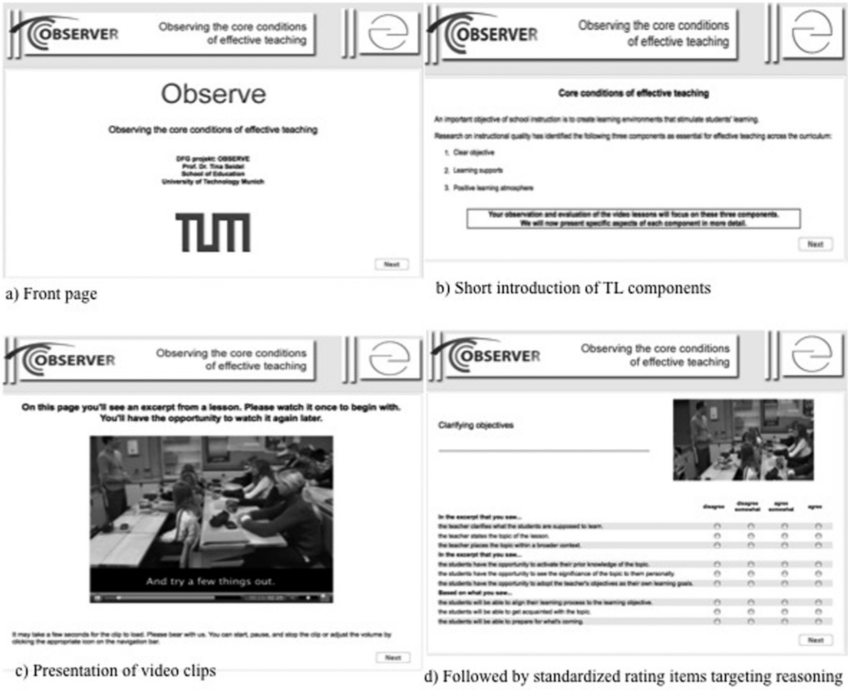

Figure 3. **The Observer instrument.**

Study 1

Participations were N = 119 preservice teachers (65.5% female). They had a mean age of 21.84 years (SD = 3.03) and were on average in their fifth semester of their initial university teacher education program (M = 4.64). The participants also showed variance across study semesters (SD = 3.89). Besides sociodemographic data, we used an open-format question to assess how many general pedagogical courses on the topic of teaching and learning the participants had attended (M = 3.98, SD = 2.80).

Study 2

Participants were N = 152 preservice teachers (59.2% female). They had a mean age of 21.57 years (SD = 1.49), were on average in the fifth semester (M = 4.97, SD = 0.18), and on average attended three courses on teaching and learning (M = 2.52, SD = 1.10). With regard to differences between both samples, the participants did not differ in age, ΔM = 0.28, SD = 0.28,

$t(269) = 0.98$, $p = .33$, $d = 0.12$; and study semester, $\Delta M = 0.33$, $SD = 0.32$, $t(269) = 1.03$, $p = .30$, $d = 0.13$. Systematic differences were found with regard to previous attendance of teaching and learning courses, $\Delta M = 1.46$, $SD = 0.25$, $t(269) = 5.90$, $p < .01$, $d = 0.72$, which can be explained by the higher variance of study semesters in the first sample.

Study 3

Participants were $N = 20$ preservice teachers (female 75%) who studied in advanced semesters (6/7) of their university teacher education program. All participants did not complete the Observer instrument before. No further information was collected.

Research Design

Study 1

To ensure that items are locally independent on the video clips, we conducted the first study with 12 video clips as "prompts" to elicit teacher knowledge. Two Observer test versions (test version A and B) were implemented, each including half of the original set of 12 video clips. Video clips were systematically varied with respect to the subject shown and the represented TL components. Additionally, video clips in the test versions were rotated to prevent order effects (see Table 2 in the Results section). Both versions were tested in the same time span. All participants were enrolled in the same teacher education program and were taught professional teacher knowledge (content knowledge, pedagogical content knowledge, and generic pedagogical knowledge). This stage in their teacher education program did not yet include systematic practical teaching experience. Preservice teachers were invited to participate in the study on a voluntary basis. They received an online link hosting a randomly assigned version of the Observer instrument (including respective rotations). The instrument had to be completed within one week. Participants randomly assigned to the two test versions (test version A: $N = 66$; test version B: $N = 53$) did not differ in age, $\Delta M = 0.32$, $SD = 0.56$, $t(117) = 1.12$, $p = .27$, $d = 0.21$; semester, $\Delta M = 0.81$, $SD = .73$, $t(115) = 0.58$, $p = .56$, $d = 0.11$; or attended courses on teaching and learning, $\Delta M = 0.58$, $SD = 0.56$, $t(117) = 1.13$, $p = .26$, $d = 0.21$.

Study 2

Based on the findings of Study 1, a final version of the Observer instrument was created. In order to test our model and assumptions about the structure of professional vision (Research Question 2), a second study including the final version of the instrument was conducted. Participants were enrolled in the same teacher education program as in Study 1. Before entering the fifth semester of their study, program participants

Table 2
Clip Characteristic in Observer Versions A (with additional rotation) and B (Study 1)

Represented Teaching and Learning (TL) Component	Clip Number	Test Version A (N = 66) Clip Position (rotation)	Discernible Example	Test Version B (N = 55) Clip Number	Discernible Example
Goal clarity	1	1; 4	.78 (.42)	7	.79 (.41)
Learning climate			.65 (.48)		.59 (.50)
Goal clarity	2	2; 5	.56 (.50)	8	.50 (.51)
Teacher support			.63 (.49)		.86 (.35)
Learning climate	3	3; 6	.76 (.43)	9	.86 (.35)
Teacher support			.87 (.34)		.84 (.37)
Goal clarity	4	4; 1	.85 (.36)	10	.89 (.31)
Learning climate			.72 (.45)		.80 (.40)
Goal clarity	5	5; 2	.55 (.50)	11	.43 (.50)
Teacher support			.75 (.44)		.80 (.40)
Learning climate	6	6; 3	.81 (.40)	12	.87 (.33)
Teacher support			.88 (.33)		.80 (.40)

Note. Discernible example: percentage participant agreement with experts.

were asked in the first session of their teaching and learning course to participate in the study. Participation was voluntary. Similar to Study 1 the participants received an online link hosting the final version of the instrument. The instrument had to be completed within one week.

Study 3

In order to answer the research question on retest reliability and learning effects, a third study was conducted. Participants were preservice teachers who were asked in teacher education program courses for voluntary participation. Preservice teachers were invited to assess their professional vision on site in a pre-post design. They were randomly assigned to the four study conditions: four groups repeated the Observer instrument with an interval of 3 (n = 5), 7 (n = 4), 14 (n = 6), and 21 days (n = 5).

Instruments

Observer Instrument

Participants in the three studies completed the Observer instrument (for details see previous section). In Study 1 two test versions (and rotations in version A) were used. In Studies 2 and 3 the final version of the instrument was completed.

Video Clips as Discernible Examples of TL Components

After observing each video clip, participants were asked to judge (yes/no answer) which TL component had been addressed in the video clip. These judgments were used to analyze whether the selected video clips were discernible examples of TL components.

Data Analysis

Video Clips as Discernible Examples of TL Components

For each video clip the research team checked which TL component was represented in the video (goal clarity: yes/no, teacher support: yes/no, learning climate: yes/no). After watching a video clip participants were also asked to check which TL component was represented according to their opinion. We then calculated the mean agreement (in form of percentages) between participants and research team. The mean agreement calculation was conducted twice, for participations of Study 1 as well as for participations of Study 2.

Testing the Model of Professional Vision

We followed the requirements of item response theory (IRT) and used Rasch models as a standard in educational test construction (i.e., Organisation for Economic Co-operation and Development, 2005). It was assumed that a measured response of participants in rating video clips with the Observer instrument is related to a latent trait (professional vision). Assuming that professional vision (Figure 1) can be measured as a continuous one-dimensional latent variable, it should be possible to distinguish qualitative levels of person abilities (Wilson, 2005). In our case of a validation, Rasch models provide important prerequisites such as local stochastic independence and one-dimensionality (Embretson & Reise, 2000). For these reasons we decided to use and test Rasch models. We are, however, aware of the fact that other models such as 2 PL models in which item discrimination for each item are freely estimated might fit data better and provide additional information.

Identical discrimination of items is a requirement of applying Rasch models in order to construct homogeneous scales and test for distinct dimensions in measurement. Based on a maximum likelihood estimation procedure, two types of parameters (item difficulty and person ability) can be estimated indepentently on the same logit scale (Moosbrugger & Hartig, 2002), producing an ideal likelihood for the data. Weighted person ability scores (mle) depend on item characteristics as well as person abilities and describe the performance in professional vision of preservice teachers in our sample.

Model-based measuring (Embretson, 1996) is a suitable approach for empirically testing the structure of professional vision. Given that we focus on a multidimensional interpretation of performance in professional vision, modeling separate, though not unrelated, dimensions of description, explanation, and prediction is a prerequisite before measuring professional vision as an overall construct (Briggs & Wilson, 2003). In this vein, we applied a three-dimensional Rasch model using the software ConQuest (Wu, Adams, & Wilson, 1997). Multidimensional analyses (Briggs & Wilson, 2003) examine the different subscales simultaneously, incorporating the responses to all items and capitalizing on the correlations between subscales (Allen & Wilson, 2006).

In order to test the structure of professional vision, a comparison between the three-dimensional model assuming description, explanation, and prediction each to be measured as distinct dimensions and the more restrictive one-dimensional model (capturing all items on one dimension), and the two-dimensional model assuming two dimensions of description and explanation/prediction (as integration) was conducted. In addition, the models were tested using two different expert reference norms (expert "hit/miss" vs. approximation "hit, close match, miss"; see Methods section). Participants' ratings were compared in terms of their agreement with the two expert reference norms: a dichotomous model was used for the strict norm of hit/miss; a partial credit model was used for the approximation reference norm (hit/close match/miss), based on a step-parameter for ordinal data. In total, six models (three professional vision models, two types of reference norms) were compared with regard to their item fit indices (Mean Square Fit Index) and scale indices (EAP/PV reliability and explained variance indicating item discrimination). In addition, the model fit of the three-dimensional model is compared to a one-dimensional and a two-dimensional model by applying the likelihood ratio test for the final deviance of the model and calculating the Bayesian Information Criterion (BIC). Figure 4 provides an overview of the scaling design. The scaling design was applied to Study 2 in which participants completed the final version of the Observer instrument.

Given that absolute fit statistics are not provided for Rasch models, we additionally conducted confirmatory factor analyses (CFA) for the six models by using the software Mplus (Muthèn & Muthèn, 1998–2010). CFA analyses for categorical variables are also referred to as item response theory analysis (Baker & Kim, 2004), though item discriminations for each item are freely estimated. By drawing on the root mean square error of approximation (RMSEA), we see indications for the general fit of data to the theoretically assumed model of professional vision.

Analyzing Repeated Measurement Effects

In order to answer the third research question, we used the R-software package nparLD (Noguchi, Gel, Brunner, & Konietschke, 2012) for

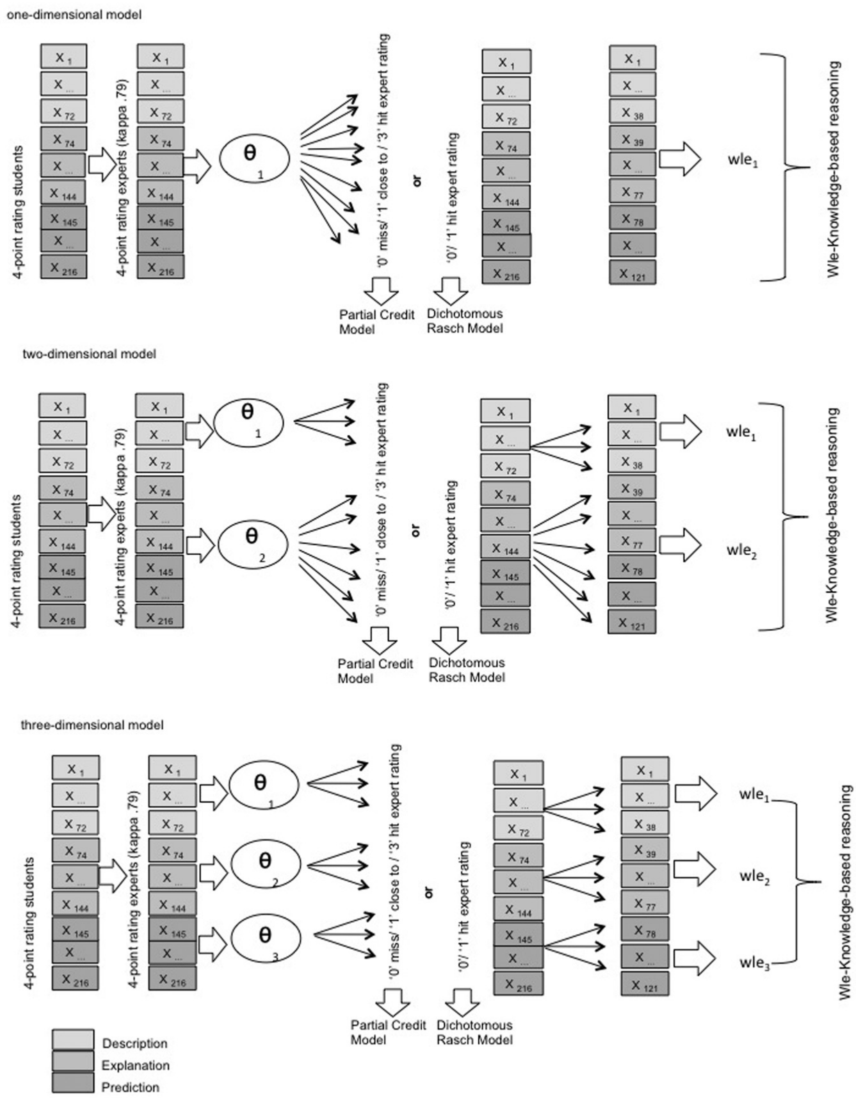

Figure 4. **Scaling design.**

nonparametric variance analysis (Time × Group) of longitudinal data. Dealing with even small sample sizes, this method allowed us to measure the relative treatment effect (RTE) based on the mean ranks.

Results

Video Clips as Discernible Examples of Teaching and Learning Components

In the development of an instrument to assess teacher professional vision using video prompts, the first objective is to select sequences that are relevant for student learning. Based on teaching effectiveness research, three TL components of goal clarity, teacher support, and learning climate were identified as relevant and video sequences representing TL components were selected. In a pilot study, expert researchers rated the video sequences as authentic and representative for the chosen TL components. In this study, participants of Studies 1 and 2 were asked to assess each observed video clip for the TL component addressed. In Tables 2 and 3 the results of these assessments ("discernible example") are shown. In Study 1 (Table 2), the participants who were randomly assigned to one of the two test versions (A/B) agreed with the predefined assessment of the video clips by the research team to a substantive amount: 66.9% of the participants in Study 1 rated the video clips in accordance with the research team for the component of goal clarity, 80.4% for teacher support, and 75.8% for learning climate.

In Study 2 (Table 3), similar assessments were found (62.0% for goal clarity, 76.0% for teacher support, 80.0% for learning climate). Thus, the participants strongly agreed that the video clips represented discernible examples of TL components. The participants had no knowledge of the assessment of the video clips through the research team.

Testing the Model of Professional Vision

In our second research question, we empirically tested with Study 2 whether a model operationalizing professional vision into three dimensions (description, explanation, and prediction) describes preservice teachers' reasoning best.

First, the psychometric properties of the 216 rating items were examined for the three-dimensional model and compared with a more restrictive one- and two-dimensional model. Items were analyzed in terms of the Mean Square Fit Index (MNSQ $\leq .75 \geq 1.30$ see: Bond & Fox, 2001). The application of these criteria resulted in an item pool of $N = 112$ items that fit the three-dimensional model. Comparing both reference norms, the MNSQ values exceeded the range for good model fit in all partial credit models, whereas the fit was within the range for the dichotomous models.

Table 3
Clip Characteristic in Final Observer Version (Study 2)

Represented Teaching and Learning (TL) Component	Clip Number	Clip Position	Discernible Example
Goal clarity	1	1	.93 (.25)
Learning climate			.70 (.46)
Goal clarity	11	2	.25 (.43)
Teacher support			.80 (.40)
Learning climate	12	3	.83 (.37)
Teacher support			.85 (.36)
Goal clarity	10	4	.92 (.27)
Learning climate			.79 (.41)
Goal clarity	2	5	.38 (.49)
Teacher support			.65 (.48)
Learning climate	3	6	.88 (.33)
Teacher support			.74 (.44)

Note. Discernible example: percentage participant agreement with experts.

To receive an exact estimation of scale indices (Rauch & Hartig, 2010), unidimensional model estimations for the proposed professional vision scales were then run in a second step of analysis. The scale indices for the six models of professional vision and expert reference norms for the final Observer instrument are given in Table 4.

A comparison of scale indices for the two expert reference norms (dichotomous, partial credit) shows for all model comparisons best indices for the strict expert norm that was used in the dichotomous model. The partial credit models show acceptable reliability scores but a low discrimination of the scales (explained variance).

Regarding the fit of the assumed three-dimensional model of professional vision, the results of the model comparisons shows best indices with good (description and explanation) to excellent (prediction) reliabilities and good item discrimination with up to $\sigma^2 = 2.14$ explained variance for the prediction scale.

In a third step, we compared the global model fit (deviance, parameter, Δdeviance, BIC) between the dichotomous three-dimensional and the one-dimensional as well as the two-dimensional model (Table 5). The likelihood ratio test shows that the three-dimensional model fits the data significantly better than the more restrictive one- and two-dimensional models. The comparably smaller BIC values for the three-dimensional model confirm these findings.

In addition, we used CFA to determine absolute fit statistics for each of the tested model. These analyses show for all models an absolute fit index

Table 4

Indices for the Professional Vision Scales Final Observer Version (*N* = 152)

	One-Dimensional Model	Two-Dimensional Model	Three-Dimensional Model	One-Dimensional Model	Two-Dimensional Model	Three-Dimensional Model
Expert Reference Norm	Dichotomous Rasch Models(1 = hit; 0 = miss)			Partial Credit Models (2 = hit; 1 = close; 0 = miss)		
Reliability	0.99			0.95		
Description		0.90	0.90		0.85	0.85
Explanation		0.90	0.91		0.99	0.94
Prediction			0.97			0.90
Variance	1.24			0.43		
Description		0.80	0.80		0.28	0.28
Explanation		1.70	1.33		0.54	0.36
Prediction			2.14			0.85

Note. EAP/PV reliability and variance values based on estimations of unidimensional models for the different professional vision scales.

Table 5

Testing the Structure of Professional Vision: Model Comparison for the Dichotomous Rasch Models for the Final Observer Version (*N* = 152)

	Deviance	Parameter	ΔDeviance (*df*)	BIC
Three-dimensional	16,874	118	—	17,466
One-dimensional	17,002	113	128**(5)	17,569
Two-dimensional	16,898	115	24**(3)	17,475

Note. The three-dimensional model constitutes the baseline for model comparison. Deviance: –2log (likelihood ratio) of model estimation; ΔDeviance: χ^2-distributed test value of the likelihood ratio test with the difference of estimated parameters in the models as degree of freedom (*df*).
**$p < .001$.

value that is considered as a good fit to the data. Thereby, the absolute fit for the three-dimensional model is slightly better (RMSEA = .04) compared to the two-dimensional model (RMSEA = .05) and one-dimensional model (RMSEA = .05).

Given these results, professional vision as assessed with our instrument can be measured best as three abilities of description, explanation, and prediction. Person ability parameters for each of the three dimensions were then estimated, resulting in one score for each dimension. Since the one-dimensional model showed excellent reliabilities, person ability parameters were also estimated for professional vision as an overall score. For further analyses, four scores were used: professional vision (overall), description, explanation, and prediction (Table 6).

Table 6
Mean Person Ability Scores and Intercorrelations for Professional Vision Scales

Scale	Professional Vision (overall)	Description	Explanation	Prediction
Professional vision (overall)	.37 (.18)			
Description	.92**	.44 (.18)		
Explanation	.95**	.82**	.31 (.18)	
Prediction	.94**	.77**	.89**	.35 (.22)

Note. Scores, representing percentage agreement with experts values, on the main diagonal (standard deviation in parentheses); bivariate Pearson correlations of person ability parameters between the scales below.
**$p < .001$ (two-tailed).

Since previous qualitative research and theoretical conceptualizations would assume that the three dimensions are highly interrelated with regard to the integration and transfer of knowledge about teaching and learning to classroom situations, we additionally calculated bivariate (Pearson) latent correlations of person ability parameters among the three dimensions and the overall score for professional vision. The findings show that description, explanation, and prediction are all closely interrelated and related to the overall score.

Repeated Measurement Effect

In addressing our third research question, we tested whether the Observer instrument provides a consistent measurement over time. Therefore, we used data of Study 3 in which preservice teachers were randomly assigned to one of four study conditions. The four conditions included the repeated completion of the final version of the instrument. The interval between retest was varied between 3 (n = 5), 7 (n = 4), 14 (n = 6), and 21 days (n = 5). In Table 7, means and standard deviations for the overall professional vision score at Time 1 and Time 2 as well as the mean ranks for all four groups are given. The nonparametric variance analysis shows no systematic differences in mean ranks between the two measurement points over the groups (F = 0.15, df = 1, p = .70) and no interaction effect for time and group (F = 0.29, df = 2.36, p = .78).

Discussion

The aim of the present study is to advance the scientific understanding of professional vision by modeling and empirically testing certain characteristics of the construct in the context of university-based teacher education and to develop and provide a contextualized and yet standardized

Table 7
Repeated Measurement Effects (professional vision overall)

	Time 1		Time 2		Time 1	Time 2
Intervals	M^a	*SD*	M^a	*SD*	Mean Rank	Mean Rank
Three days (n = 5)	.28	.07	.28	.09	14.20	16.50
One week (n = 4)	.40	.08	.39	.10	30.88	28.13
Two weeks (n = 6)	.34	.05	.34	.06	23.00	22.50
Three weeks (n = 5)	.30	.07	.29	.08	16.90	14.60

[a]Scores represent percentage agreement with experts.

instrument that—in the long term—could be used as a formative assessment instrument. The findings of our research provide some evidence that the video clips used in our instrument served the purpose to represent discernible examples of relevant teaching and learning components (Research Question 1). Furthermore, the findings indicate that the developed instrument functioned in the proposed way by assessing professional vision with three subaspects of description, explanation, and prediction (Research Question 2). Finally, our data provide some evidence that without further intervention the instrument measures professional vision stable over time (Research Question 3).

Video Clips as Discernible Examples of Teaching and Learning Components

Given the contextualized and situated nature of teacher knowledge (Borko, 2004), we used videotaped classroom sequences as prompts to elicit teacher knowledge. A study of Kersting (2008) served as an example in which video clips were used as prompts and combined with open questions tackling teacher noticing and reasoning. However, the situations shown in the video clips had a rather general function without further definition of what teachers are supposed to notice and to reason upon. This approach is also taken by a number of other studies in the field of preservice teacher research (e.g., Santagata & Angelici, 2010; Santagata, Zannoni, & Stigler, 2007; Star & Strickland, 2008).

In our study we expanded this approach by using findings of teaching effectiveness research as a basis for predefining what preservice teachers might be able to notice and how they might reason about these noticed components. This is a major contribution to the field in which judgments of teaching and learning situations are often regarded as being too normative to address the complexity and "art" of classroom teaching and learning (cf. Grossman & McDonald, 2008). Teaching effectiveness research, however, has provided quite stable findings with regard to relevant teaching and learning components over decades (Fraser et al., 1987; Hattie, 2009; Seidel & Shavelson, 2007). These findings are helpful for teacher research since they provide "lenses" to systematically analyze and explain teaching and learning situations, combined with empirical evidence about possible effects on student learning. In taking this approach in our study for the first time, three teaching and learning components were identified (goal clarity, teacher support, learning climate) and video clips representing TL components were selected. In our first research question we tested whether the selected video clips were equally regarded by the participants in our studies as respective examples of TL components. Thereby, we could show a high agreement with the pre-assignment through the research team, while the participants have not been informed about this pre-assignment. A pilot study with external experts in teaching effectiveness research resulted in the similar finding (Seidel et al., 2010a). In focusing on the three identified TL components, we are aware that the interpretations being drawn from the use of the instrument have to be restricted to this context (Kane, 1994) and that we have not captured all relevant teaching and learning components that occur in the complex environment of classrooms. Based on the teaching and learning model used in our study (Seidel & Shavelson, 2007), a substantial body of relevant components were addressed. So far, our results support the idea that TL components are discernible in video clips. Based on these findings, extensions of the instrument can be accomplished regarding the inclusion of additional video clips as well as expanding the instrument according to our model of teaching and learning.

Testing the Structure of Some Aspects of Professional Vision in Preservice Teachers

The further significant contribution of the presented study is to integrate previous findings of qualitative research into a model of professional vision and to test the model using a quantitative empirical approach (second research question). Multidimensional statistical models based on item response theory were used to test three models with regard to their fit indices (Briggs & Wilson, 2003): a one-dimensional model assuming one ability of professional vision, a two-dimensional model differentiating between descriptions and explanations/predictions (integration), and a three-dimensional model with the three abilities of description, explanation, and prediction as distinctive dimensions. The results can be taken as support for the idea that professional vision (as captured with the developed instrument) can be differentiated along the three assumed abilities. We acknowledge that the current IRT software does not allow providing absolute fit statistics for Rasch models within the given data set. We decided, however, to use Rasch models because by using equal discrimination parameters we valued the strictness of constructing homogenous scales for model comparisons and based our statistical procedures on the standards common in this field (i.e., Organisation for Economic Co-operation and Development, 2005). To further test our assumptions, we additionally conducted confirmatory factor analyses in which item discriminations for each item are freely estimated. These analyses help to investigate the absolute fit between theoretically assumed models and empirical data. In using RMSEA as an indicator, we also see indication for a good fit.

Model comparisons for our empirical study showed best scale indices for the three-dimensional model. The three-dimensional model resulted in good reliabilities for each of the three abilities and in an excellent explanation of variance for the final test version of the instrument. In accordance with our assumptions, our data can be taken as support for the idea that the three abilities are highly interrelated and that they substantially relate to the overall construct of professional vision. So, with the Observer measurement—including expert responses in the item development process—we provide construct-related evidence for the validity of professional vision scores (Koeppen et al., 2008). We acknowledge that multiple lines of evidence are required for further validating our instrument, for example, by investigating different samples ranging in expertise or relating our instrument to other criterion-related measures. A first attempt in this direction has been accomplished by comparing the results of our quantitative measure with qualitative analyses of written responses (to an open question format) of participants using one of the video clips of the instrument. The results of this study show significant interrelations between the level of reasoning as shown by qualitative analysis and the scores achieved in our quantitative measure (Schäfer & Seidel, in press). For further research, it should be studied in more detail whether the ability to describe, explain, and predict classroom situations requires different processes and knowledge structures. Further studies in the context of our project indicate that this might be the case (Seidel, Blomberg, & Renkl, 2013). Interventions in which conditions of knowledge acquisition are varied might be helpful to learn more about the specific processes and knowledge structures involved.

Stability of Measurement Over Time

Furthermore, it was tested to what extent learning effects occurred while completing the instrument (third research question). Retest reliability was tested with intervals ranging between 3 and 21 days. Due the small sample size, we used the R-package nparLD for nonparametric variance analysis of longitudinal data that has been shown as robust method even for small sample size (Noguchi et al., 2012). Independent of the time between both test completions, the retest scores closely matched pretest scores, indicating some stability over time. With regard to the fact that the Observer instrument does not provide any form of feedback but in turn, adequate feedback is seen as essential for efficient learning and performance improvement (Ericsson et al., 1993), the finding seems to underpin the claim that no learning effects occur through its repeated processing. It can be taken as support for the formative use of the Observer instrument in the evaluation of teacher education programs. In our own research project, for example, we used the Observer instrument to capture changes and development during the acquisition of professional knowledge in teacher education programs. By using the Observer instrument as a pre- and posttest measure, we could show that preservice teachers positively developed professional vision while attending university courses on principles of teaching and learning (Stürmer, Könings, & Seidel, 2013) as well as during a theory practice term (Stürmer, Seidel, & Schäfer, 2013).

In teacher education research a call is given to describe "job-oriented" professional demands and to develop valid instruments to capture the specific professional competencies as they are acquired in various stages of teacher education (Darling-Hammond, 2006, 2010). In our study we focused on professional vision as a key element of teacher expertise (van Es & Sherin, 2002) that has been shown to develop during a professional teaching career (Berliner, 2001; Putnam & Borko, 2000). The competence to observe and interpret classroom situations is one of the first major "job demands." It is especially addressed in the early and initial phases of university teacher education programs. In this phase, future teachers are acquiring relevant knowledge about principles of teaching and learning. To acquire this knowledge in a way that preservice teachers can transfer it to actual classroom situations is of high importance (Grossman & McDonald, 2008), especially since it may significantly influence the continuous professional development of candidate teachers (Blomberg, Renkl, Sherin, Borko, & Seidel, 2013; Brouwer, 2010). Preservice teachers who have developed professional vision are more likely to use this knowledge when observing lessons in the context of sitting in on classes. Furthermore, the competence to analyze one's own teaching practice according to the three facets represents a necessary prerequisite for lifelong learning (Santagata & Angelici, 2010). If teachers, for example, are able to describe their own practice without further judgment, explain observed practice with regard to concepts of teaching, and use evidence-based knowledge in order to draw conclusions with regard to effects on student learning, an important precondition for continuously redefining teaching practice would be achieved.

Especially when it comes to designing learning environments in university-based education, we see a valuable contribution of our study. The Observer has the potential to support instruction that leads preservice teachers toward developing professional vision even before they enter a classroom for the first time. The video-based approach allows students to be closer to authentic teaching situations while at the same time not pressuring candidates to act in a classroom (Darling-Hammond, 2010). Thus, the use of video-based approaches and instructional conditions that foster professional vision in university-based teacher education can be seen as one form of providing preservice teachers with "approximations of practice" (Grossman & McDonald, 2008).

Conclusion

Given the results of this study, we argue that professional vision as an element of teacher expertise can be assessed through the use of a standardized, yet contextualized, measurement approach. This study is the first to investigate the qualitatively and theoretically derived structure of professional vision empirically and to use a quantitative approach. It advances the scientific understanding of the concept and provides an example of how theoretically informed development and testing can be applied to an assessment instrument for use in the field of teacher education.

Note

The Observe project (SE 1397/2-2) is funded by the German Research Foundation as part of the priority research program "Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes." We would like to thank the teacher candidates who participated in this study, Susannah Goss and Alison Lowry for editing the manuscript, as well as Dr. Geraldine Blomberg for supporting us in the procedure of the study.

References

Allen, D., & Wilson, M. (2006). Introducing multidimensional item response modeling in health behavior and health education research. *Health Education Research, Theory & Practice*, *21*(Suppl. 1), 73–84. doi:10.1093/her/cy1086

Baker, F. B., & Kim, S. (2004). *Item response theory. Parameter estimation techniques* (2nd ed.). New York, NY: Marcel Dekker, Inc.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. doi:10.3102/0002831209345157

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

Blomberg, G., Renkl, A., Sherin, M. G., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in preservice teacher education. *Journal of Educational Research Online*, *5*(1), 3–33.

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

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model*. Mahwah, NJ: Erlbaum.

Borko, H. (2004). Professional development and teacher learning: mapping the terrain. *Educational Researcher*, *33*(8), 3–15. doi:10.3102/0013189X033008003
Borko, H., Jacobs, J. K., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417–436.
Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, *26*(4), 473–498. doi:10.3102/00028312026004473
Boshuizen, H. P. A., Schmidt, H. G., Custers, E. J. F. M., & van de Wiel, M. W. (1995). Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction*, *5*, 269–289. doi:10.1016/0959-4752(95)00019-4
Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, *4*(1), 87–100.
Brophy, J. (Ed.). (2004). *Using video in teacher education*. Amsterdam: Elsevier.
Brouwer, N. (2010). Determining long term effects of teacher education. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 7, pp. 503–510). Oxford: Elsevier.
Buff, A., Reusser, K., Rakoczy, K., & Pauli, C. (2011). Activating positive affective experiences in the classroom: "Nice to have" or something more? *Learning and Instruction*, *21*(3), 452–466. doi:10.1016/j.learninstruc.2010.07.008
Carter, K., Sabers, D., Cushing, K., Pinnegar, P., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice and postulant teachers. *Teaching and Teacher Education*, *3*, 147–157. doi:10.1016/0742-051X(87)90015-1
Cochran-Smith, M. (2003). Assessing assessment in teacher education. *Journal of Teacher Education*, *54*(3), 187–191.
Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: LEA Publishers.
Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, *57*(2), 120–138. doi:10.1177/0022487105283796
Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, *61*(1-2), 35–47. doi:10.1177/0022487109348024
Darling-Hammond, L., & Bransford, J. D. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Franscisco, CA: Jossey-Bass.
Deci, E. L., & Ryan, R. M. (Eds.). (2002). *The handbook of self-determination research*. Rochester, NY: University of Rochester Press.
Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*(4), 341–349. doi:10.1037/1040-3590.8.4.341
Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acqusisition of expert performance. *Psychological Review*, *100*, 363–406.
Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, *11*, 145–252.
Goldman, R., Pea, R., Barron, B., & Denny, S. J. (Eds.). (2007). *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum.

Goldstein, G., & Hersen, M. (2000). *Handbook of psychological assessment*. Oxford: Elsevier.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633. doi:10.1525/aa.1994.96.3.02a00100

Grossman, P., & McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205. doi:10.3102/0002831207312906

Hammerness, K., Darling-Hammond, L., & Shulman, L. S. (2002). Toward expert thinking: How curriculum case-writing prompts the development of theory-based professional knowledge in student teachers. *Teacher Education*, *13*(2), 221–245. doi:10.1080/1047621022000007594

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. doi:10.3102/00028312042002371

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539. doi:10.1007/s10648-007-9054-3

Kane, M. T. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation & the Health Professions*, *17*(2), 133–159. doi:10.1177/016327879401700202

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. doi:10.1177/0013164407313369

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie-Journal of Psychology*, *216*(2), 61–73. doi:10.1027/0044-3409.216.2.61

Koster, B., Brekelmans, M., Korthagen, F., & Wubbels, T. (2005). Quality requirements for teacher educators. *Teaching and Teacher Education*, *21*(2), 157–176.

Kunter, M., Baumert, J., & Koller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, *17*(5), 494–509. doi:10.1016/j.learninstruc.2007.09.002

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. doi:10.1016/j.learninstruc.2008.11.001

Muthèn, L. K., & Muthèn, B.O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthèn & Muthèn.

Moosbrugger, H., & Hartig, J. (2002). Factor analysis in personality research: Some artefacts and their consequences for psychological assessment. *Psychologische Beiträge*, *44*(1), 136–158.

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*(12), 1–23.

Organisation for Economic Co-operation and Development. (2005). *PISA 2003 technical report*. Paris: Author.

Oser, F., Heinzer, S., & Salzmann, P. (2010). The measurement of the qualtity of professional competence profiles of teachers. *Unterrichtswissenschaft*, *38*(1), 5–29.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*(1), 4–15.

Rauch, P., & Hartig, J. (2010). Multiple-choice versus open-ended responses formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, *52*(4), 354–379.

Resnick, L. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Reusser, K. (2005–2009). *Videoportal* [Training videos]. Retrieved from http://www.didac.uzh.ch/videoportal/

Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education*, *61*(4), 339–349. doi:10.1177/0022487110369555

Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in preservice teacher education: an empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, *10*(2), 123–140. doi:10.1007/s10857-007-9029-9.

Schäfer, S., & Seidel, T. (in press). Noticing and reasoning of teaching and learning components by preservice teachers. *Journal of Educational Research Online*.

Seidel, T., Blomberg, G., & Stürmer, K. (2010a). "Observer"—Validation of a video-based instrument for measuring the perception of professional education. *Zeitschrift für Pädagogik*, *56*, 296–306.

Seidel, T., Blomberg, G., & Stürmer, K. (2010b). Observer: video-based tool to diagnose teachers' professional vision. [Software]. Unpublished instrument, retrieved from http://ww3.unipark.de/uc/observer_engl/demo/kv/

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, *34*, 56–65. doi:http://dx.doi.org/10.1016/j.tate.2013.03.004

Seidel, T., & Prenzel, M. (2007). How teachers perceive lessons—Assessing educational competencies by means of videos. *Zeitschrift für Erziehungswissenschaft*, *10*, 201–216.

Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study: Technical report of the IPN Video Study*. Münster, Germany: Waxmann.

Seidel, T., Rimmele, R., & Prenzel, M. (2003). Opportunities for learning motivation in classroom discourse—Combination of video analysis and student questionnaires. *Unterrichtswissenschaft*, *31*(2), 142–165.

Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction*, *15*(6), 539–556, doi:10.1016/j.learninstruc.2005.08.004

Seidel, T., & Shavelson, R. J. (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*(4), 454–499. doi:10.3102/0034654307310317

Shavelson, R. J. (1991). Generalizability theory and military performance measurements: I. Individual performance. In A. K. Wigdor & B. F. Green (Eds.), *Performance assessment for the workplace*. Washington, DC: National Academies Press.

Sherin, M. G., Jacobs, V. R., & Randolph, P. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York, NY: Routledge.

Sherin, M. G., & van Es, E. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, *60*, 20–37.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

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. doi:10.1007/s10857-007-9063-7

Stokking, K., Leenders, F., de Jong, J., & van Tartwijk, J. (2003). From student to teacher: Reducing practice shock and early dropout in the teaching profession. *European Journal of Teacher Education*, *26*, 329–350.

Stürmer, K., Könings, K. D., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *British Journal of Educational Psychology*, *83*(3), 467-483. doi:10.1111/j.2044-8279.2012.02075.x

Stürmer, K., Seidel, T., & Schäfer, S. (2013). Preservice teachers' professional vision changes following practical experience: a video-based approach in university-based teacher education. *Gruppendynamik und Organisationsberatung*, *44*(3), 339-355. doi:10.1007/s11612-013-0216-0

van Es, E. (2009). Participants' roles in the context of a video club. *Journal of the Learning Sciences*, *18*(1), 100–137.

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

van Es, E., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, *24*(2), 244–276.

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*(4), 952–969. doi:10.1037/a0025125

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89–122.

Wilson, M. (2005). *Constructing measures. An item responses modeling approach*. Mahwah, NJ: Lawrence Erlbaum.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software*. Camberwell: Australian Council for Education Research.

Manuscript received January 11, 2012
Final revision received October 17, 2013
Accepted November 5, 2013