Article

Assessing Biology Pre-Service Teachers' Professional Vision of Teaching Scientific Inquiry

Friederike Vogt * and **Philipp Schmiemann**

Biology Education, University of Duisburg-Essen, 45117 Essen, Germany; philipp.schmiemann@uni-due.de
* Correspondence: friederike.vogt@uni-due.de

check for
updates

Abstract: Professional vision is a key ability in the professional development of pre- and in-service teachers as it determines how professionals perceive and interpret situations. The aim of this study was to conceptualize an instrument for professional vision focusing on formative assessment in the context of scientific inquiry. This focus is highly valuable, since formative assessment contributes to the quality of science teaching and learning. The four-dimensionality of the construct of professional vision with its abilities (perception, description, explanation, and prediction) was confirmed by means of our text-vignette-based instrument. The professional vision of pre-service teachers ($N = 80$) was fostered in training, involving a seminar phase and a teaching phase in an out-of-school laboratory. In a pre-post design significant interaction effects of groups (training vs. comparison group ($N = 39$)) and time for the ability description ($F(1,117) = 29.14$ $p < 0.001$) and prediction ($F(1,117) = 14.81$ $p < 0.001$) were found, indicating the sensitivity of the instrument. Our instrument allows the assessment of the abilities description and prediction. The scales for the abilities perception and explanation need further refinements. Nonetheless, our instrument could be a starting point to further investigate professional vision in science contexts as it incorporates the essential key features such as a situated approach.

Keywords: pre-service teachers; professional development; professional vision; scientific inquiry; test instrument

1. Introduction

Among other aspects, previous research regarding science teachers' professional development focused on knowledge and beliefs of teachers. Researchers focused on Shulman's [1] conceptualization of knowledge relevant for teaching, by investigating the development and relation of pre- and in-service biology teachers' content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK) [2–6]. These types of knowledge are considered to affect teaching and student learning [7,8]. In addition, teachers' beliefs are considered a major factor affecting the strategy of teaching science subjects and an important aspect to be addressed in professional development programs [9–14]. The other main area of research in the field of science teachers' professional development focused on classroom practices of pre- and in-service teachers [15–18]. To develop a more comprehensive understanding of teachers' competence of professional development, Blömeke, Gustafsson, and Shavelson [19] conceptualized a model to bridge the gap between knowledge, beliefs, and in-classroom practice.

1.1. Professional Vision as a Part of Professional Development

Blömeke et al. [19] presented a model of professional development, presenting competence as a continuum stretching from dispositions (e.g., self-efficacy and beliefs) and teachers' knowledge (e.g., CK, PCK, PK) via situation-specific skills to actual teaching performance in the classroom

(Figure 1). Situation-specific skills are composed of the core abilities of professional vision and decision-making. Following the definition of Seidel et al. [20], professional vision determines how professionals perceive and interpret classroom situations. Professional vision involves two processes, perception and interpretation [20–22]. Perception describes the process of paying attention to an event relevant for the teaching and learning activity [23,24]. This process is also termed selective attention or noticing [20,21,25]. The process of interpretation is subdivided into three abilities: (1) description, the ability to differentiate relevant aspects of a perceived event; (2) explanation, the ability to use knowledge to reason about the event; and (3) prediction, the ability to extrapolate the consequences of the event regarding the students' learning [22,26,27]. The process of interpretation is also called knowledge-based reasoning [20]. Professional vision can be seen as an agent between knowledge (e.g., CK, PCK, and PK) and practical teaching knowledge, that is, how to act in the process of teaching [28,29]. It is a knowledge-based process [24,30,31] and is also referred to as integrated teacher knowledge [23,32] with the objective of adequately supporting the learning process.
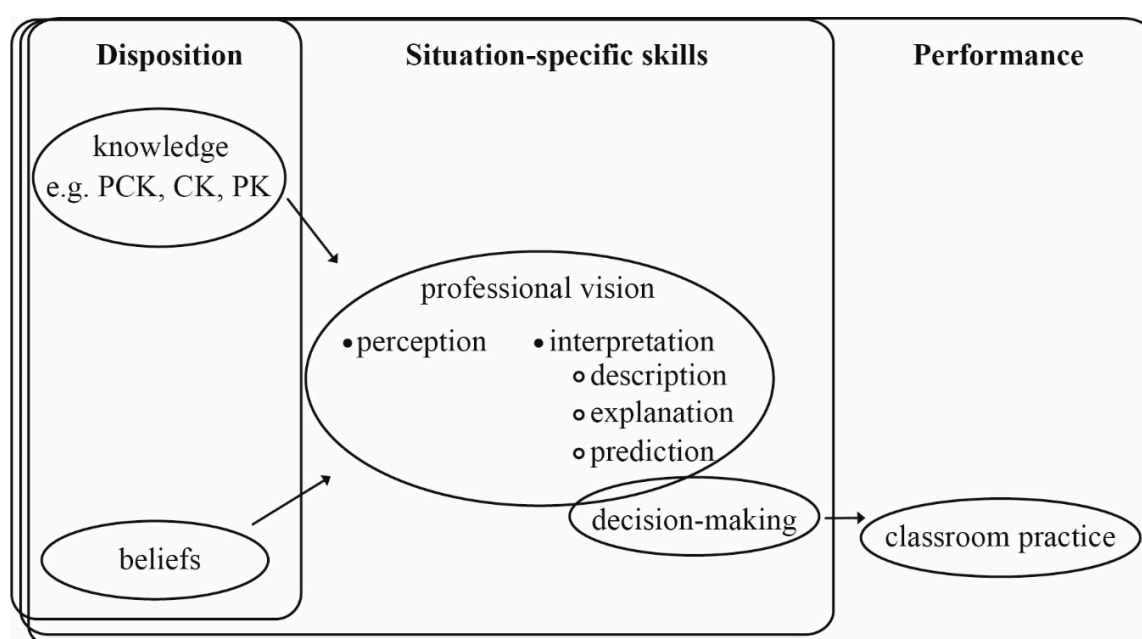

Figure 1. Model of competences of professional development based on Blömeke et al. [19].

Professional vision does not seem to be general ability but has to be reconstructed and reconsidered in the particular teaching situation. Empirical studies have demonstrated that professional vision of classroom management and content-specific learning support in science education lessons are two separate constructs [33]. Further, findings have shown that professional vision is not only content-specific but also topic-specific [34,35]. Sunder et al. [35] demonstrated this topic specificity of professional vision by means of an intervention study. The professional vision of the intervention group increased regarding the fostered ability—in this case, specific learning support concerning the topic of floating and sinking—in contrast to the specific learning support of the non-fostered topic.

1.2. Assessment of Professional Vision

Primarily qualitative studies have been conducted in the field of professional vision [21,22,36–38], giving valuable insights into the structure and procedures of professional vision. The Observer Tool [23] and the instrument developed by Möller, Steffensky, Meschede, and Wolters [39] to test the professional vision of learning support in the context of science education in primary schools are two of the few instruments that measure professional vision with a quantitative approach. A relatively new approach

to assess teachers' professional vision is by measuring their eye movements directly with the help of eye-tracking technologies [40–42].

When measuring professional vision, a situated assessment approach is needed because professional vision is the adaptation of knowledge in a specific teaching and learning situation. Therefore, open questions or closed-rating items are combined with video vignettes [23,39] or text vignettes [43,44] that serve as the key stimuli for the analyses of teaching and learning situations. As professional vision has to be reconstructed and reconsidered in the particular teaching situation, instruments used for its measurement must be aligned with the particular context that is of interest.

1.3. Training Fostering Professional Vision

Empirical studies indicate a positive relation between teachers' quality of professional vision and students' performance in mathematics [45] and science [46]. Additionally, the ability of teachers to register interactions relates to the quality of the teachers' actions in the related situations [47]. Therefore, it is important to foster the professional vision of pre- and in-service teachers. Training focusing on professional vision of pre- [35,48,49] and in-service teachers [22,46] has already shown to be successful. Some forms of training focus on general pedagogical aspects of teaching [36,48,50] such as goal clarity, teacher support, and learning climate [31]. The majority of studies have centered on the development of subject-specific educational teaching aspects [22,38,46] like analyzing the teaching of mathematic procedures [51] or providing learning support for physical phenomena [35].

An overarching concept of these forms of training is the implementation of videotaped teaching and learning sessions. The success of video-based professional development programs such as the STeLLA program highlights the valuable aspects of using videos and their effectiveness on science teaching and learning [52,53]. The way of implementing videos in professional development programs is key as participant-centered discussions can foster the professional vision of the participants [54].

Specifically, pre-service teachers benefit from the realistic yet less complex presentations of practical teaching that enable situational and multifaceted analyses of teaching and learning processes without the proximate pressure to act [36]. Since these videos offer a broad spectrum of opportunities to reflect and discuss pedagogical and subject educational aspects, the process of interlinking declarative, case-related, and strategic knowledge can be supported [55]. The types of videos available for analysis range from best-practice reinforcement to daily teaching methods. They are further differentiated by their authenticity (e.g., staged or real teaching) [56] as well as the variety of role players acting in the videos, for example, an unknown teacher, peers, or the participants themselves filmed during their own teaching activities [57]. Through repetitive use of these videos, the capabilities to observe, identify, and interpret can be improved [34,55,58]. Pre-service and in-service teachers highlight the importance of being able to observe effective STEM lessons. Furthermore, in-service teachers stress the relevance of analyzing videos of experienced teachers teaching STEM lessons [59]. Current research indicates that combination of analyzing videos from own teaching sessions, from peers, and from unknown teachers fosters the professional vision of pre-service teachers best [60]. Additionally, pre-service teachers' professional vision is fostered by offering video-based feedback [61].

1.4. Challenges in the Teaching and Learning of Scientific Inquiry

Professional vision can only be considered a part of teachers' professional expertise if the aspect in question is relevant for the students' learning process [21]. Therefore, we focus on professional vision of formative assessment in the specific context of scientific inquiry. Learning about scientific inquiry as well as acquiring the necessary inquiry skills form part of the educational standards in many countries [62–65]. However, students have difficulties in understanding and conducting scientific inquiry as they find the required processes and its logic challenging [66]. Typical difficulties of students conducting experiments concern all steps of scientific inquiry [67], e.g., students have difficulties formulating a relevant or even any hypothesis [68,69], have weak strategies for controlling variables [70], and have poor skills regarding data analysis [71].

Arnold, Kremer, and Mayer [72] advocate that procedural knowledge and procedural understanding will improve students' understanding of scientific inquiry. To foster the procedural understanding of scientific inquiry, inquiry-based learning is a promising approach [73]. A suitable method to teach inquiry-based learning is guided-inquiry teaching [74,75], as learning inquiry skills in a guided setting help students to overcome high cognitive challenges of open inquiry learning [76]. By practicing guided inquiry, teachers are confronted with various decision-making processes [19] such as when and how to provide support to their students [74]. To address this challenge, conducting formative assessment can help to diagnose students' prior knowledge and understanding, and to facilitate their learning processes [77,78].

1.5. The Role of Formative Assessment for Science Teaching

Formative assessment is the continuous diagnosis of the individual learning progress and the continuous response to promote learning [78–80]. Meta-analyses have indicated the general importance of formative assessment for student learning [81]. Formative assessment is also an important prerequisite for successful learning in science education [82–84]. The practice of formative assessment depends on the specific subject [78,85,86]. In-class performance of formative assessment is challenging for teachers in math and science education [86,87]. To address this challenge, formative assessment can be fostered in pre- and in-service teachers [88]. By investigating the quality of formative assessment, Furtak, Ruiz-Primo, and Bakeman [89] were able to identify four categories of teacher response quality in science teaching: (1) evaluative responses such as judging students' contributions and providing longer content-specific remarks; (2) neutral responses, including reactions that do not help students evaluate their own contributions; (3) leading responses, for example, prompts that lead to very short and oblivious students' answers; and (4) pushing responses that comprise impulses that activate students' own thinking.

1.6. Aim of the Study

Following the demands to model competences of teachers' professional development as a continuum development [19], we focused on the professional vision of pre-service teachers. As professional vision is topic-specific [34,35,48] and thus has to be reconstructed and reconsidered in the particular teaching situation, training must also focus on the particular teaching situation to foster the specific aspects of professional vision that are of interest. We decided to combine the important yet challenging field of formative assessment for teachers [78,82] with the concept of scientific inquiry—also particularly challenging for students in science education [68–71]—as a valuable point of focus for the professional development of pre-service teachers.

Our major objective was to conceptualize a test instrument enabling us to measure changes in pre-service teachers' professional vision. To evaluate the instrument we focused on three aspects: (1) dimensionality and reliability, (2) scoring of the participant's answers, and (3) sensitivity. These aspects are reflected in our research questions (RQs):

1. Dimensionality: Based on the theoretical background, we assumed a four-dimensional structure of professional vision. Hence, we explored in RQ 1: To what extent does the empirical data collected with our instrument fit this theoretically described structure of professional vision?
2. Scoring: Different expert reference norms have been used to score participants' answers in previous research. Hence, we aimed to answer the following as RQ 2: Is the use of a strict (dichotomous) or less strict (partial credit) expert reference norm more suitable?
3. Sensitivity: As any suitable measurement instrument should be able to detect changes, RQ 3 was whether our instrument is sensitive enough to measure changes of professional vision.

2. Materials and Methods

2.1. Designing a Test Instrument

We developed a test instrument to assess professional vision regarding formative assessment in the context of scientific inquiry. It is composed of text-vignette-associated items containing statements focusing on teacher response qualities within the context of students conducting experiments. A Likert scale answer was predominantly used.

2.1.1. Development of the Text Vignettes

To comply with the situated assessment approach of measuring professional vision, authentic video materials presenting microteaching situations [90], in which biology pre-service teachers supported the learning process of students conducting experiments, were recorded. These videos served as a basis for the development of the text vignettes. They were screened for passages showing students' difficulties when working on the hypothesis as a part of scientific inquiry. Three different video vignettes were selected. They differed in the type of students' difficulty when working on the hypothesis as well as in the level of response qualities of the pre-service teachers' responses. The video vignettes were transcribed and supplemented with notes about physical actions in the video, for example, when a pre-service teacher pointed at something. Thus, the final text vignettes were coherent for the reader.

2.1.2. Development of the Items

Items were developed based on the theoretical structure of professional vision (perception and interpretation: description, explanation, and prediction). The underlying system for all items was the formative assessment regarding scientific inquiry. The items interlinked the four levels of teacher response quality [89] with ways to support procedural knowledge and procedural understanding [72] systematically. This approach resulted in a set of 36 rating items per vignette. The process of perception was covered with a dichotomous item format (yes/no), whereas a four-point Likert scale (1 (disagree) to 4 (agree)) was used for items concerning interpretation. In Table 1, exemplary items for the ability of professional vision are given.

2.1.3. Scoring

Following the procedure of quantitative research in the field of professional vision, an expert rating was conducted to establish criterion-referenced norms to analyze participants' responses in an objective manner [23,39,91]. Three expert researchers independently rated all items in connection with the text vignettes according to their own professional vision. The experts were all educational researchers in biology education and had teaching experience at secondary schools. This rating showed an excellent consistency with $ICC_{unjust} = 0.86$ [92,93], indicating that the responses to the items were unambiguous and discernable. In cases of disagreement, consensus validation was performed. The participants' responses were compared to the expert rating. As the strictness of allocating points differs when using expert ratings as a criterion-referenced norm [23,35], two procedures to calculate the agreement with the experts were conducted. The following points were allocated in the strict approach: 1 (hit expert rating) and 0 (miss expert rating), whereas in the less strict approach we allocated 2 (hit expert rating), 1 (correct direction on the scale), and 0 (missed expert rating).

Table 1. Examples of items concerning the abilities of professional vision (translation from original language by the authors).

	To What Extend Are the Following Stimuli Present in the Text Vignette You Just Read: Please Choose A Plausible Answer for Each Aspect:	Answer Format
perception	"Are responses considering the clarification of what a hypothesis is present?"	Dichotomous (yes/no)
description	"The teacher explains what a hypothesis is" "The teacher explains why a hypothesis is essential in an experiment" "The teacher explains how a hypothesis is phrased"	Four-point Likert scale 1 (disagree) to 4 (agree)
explanation	"The teacher supports the understanding of what a hypothesis is." "The teacher supports the understanding regarding the hypothesis' function within the process of experimenting." "The teacher supports the correct linguistic formulation of a hypothesis."	Four-point Likert scale 1 (disagree) to 4 (agree)
prediction	"The students can transfer the hypothesis' function to other experiments." "The students are able to understand what the hypothesis within further experiments is." "The students can utilize the preferred 'If … , then … ' structure of formulating a hypothesis for further experiments."	Four-point Likert scale 1 (disagree) to 4 (agree)

The items interlink the level of teacher response quality (pushing: e.g., explain) with a certain method of facilitating the learning of scientific inquiry (procedural knowledge: e.g., what).

2.2. Structure of the Training

To assess the sensitivity of our instrument (RQ 3) we needed training to increase the professional vision regarding formative assessment in context of scientific inquiry to be measured by our test. The training consisted of two major phases, a seminar phase and a teaching phase. The core activities to foster professional vision in the seminar phase are based on the analyses of teacher–student interaction by means of both text and video vignettes, which focus on typical students' difficulties in the context of scientific inquiry [68,69] and teacher responses to support the learning process (Figure 2). In the teaching phase, one pre-service teacher attended to the learning process of the same two to three students experimenting together. During the full-day course in the out-of-school laboratory, the students carried out three experiments focusing on the topic "adaptation of animals to their habitats" [94] (p. 59) that were presented as "learning kit experiments" [95,96] (p. 57). Text vignettes used in the test instrument were not identical to the vignettes used during the training to avoid memorization effects. Fostering professional vision by making use of video vignettes was shown to be an effective approach [25,34,35,46,48–51]. The processes of practical teaching were incorporated in the teaching phase by participants supporting the learning processes of students in our out-of-school laboratory (for more details regarding the training, see [97]).
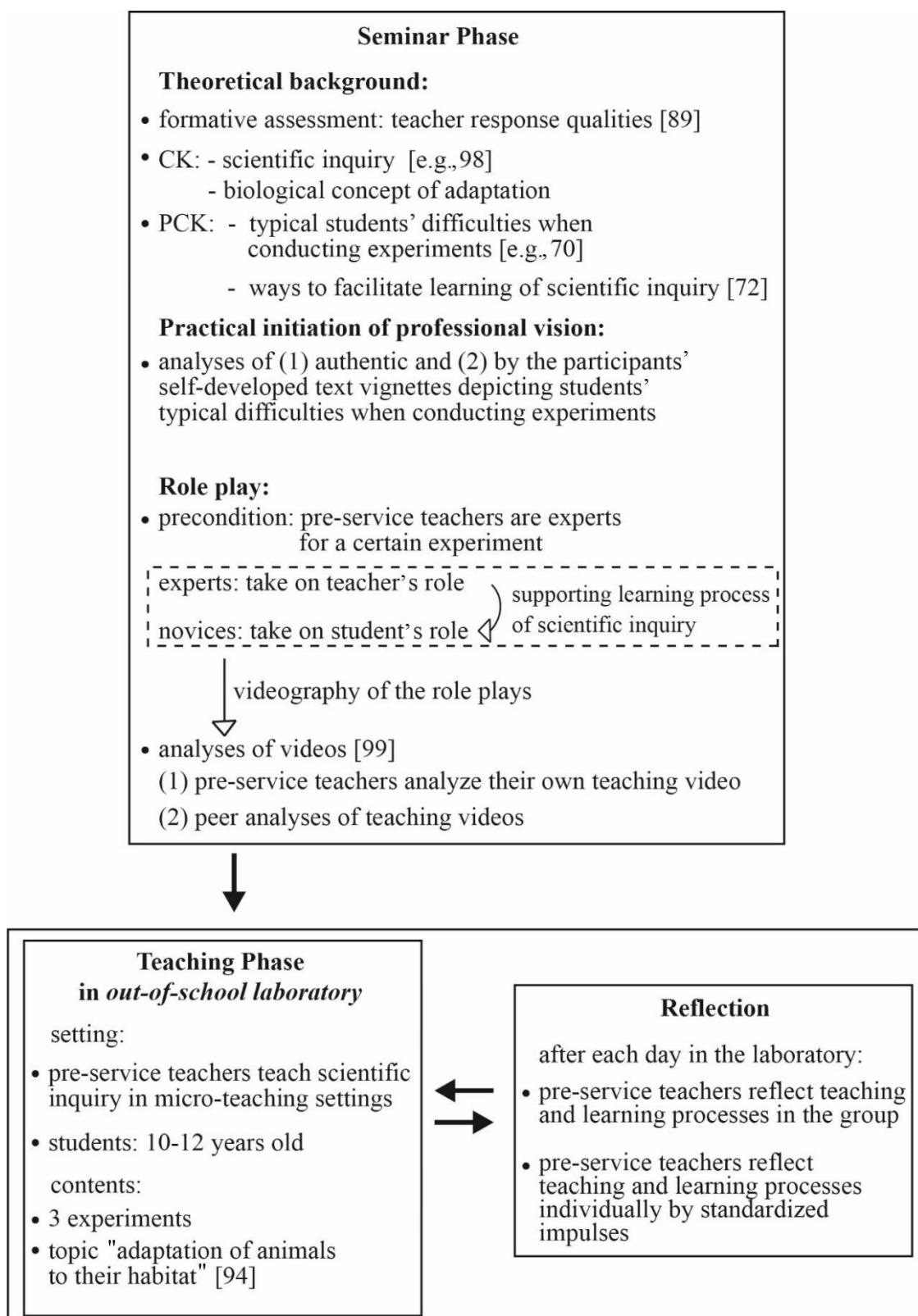
Figure 2. Structure of the training promoting professional vision regarding formative assessment in context of scientific inquiry [70,72,89,94,98,99].

2.3. Participants and Research Design

We used a quasi-experimental research design incorporating a training group and a comparison group. The training group comprised three cohorts (cohort 1: $n = 29$; cohort 2: $n = 30$; cohort 3: $n = 21$) resulting in a total of $N = 80$ biology pre-service teachers as participants (68% female). They had a mean age of 22.5 years ($SD = 2.2$) and were on average in their fifth semester of the university teacher educational program ($M = 5.0$; $SD = 1.2$). The training was divided into a seminar phase (duration: 7 consecutive days with 5 h per day) and a teaching phase in our out-of-school laboratory (duration: 5 consecutive days with 7 h per day) (Figure 3). The teaching phase was held approximately three weeks after completion of the seminar phase. The participants completed the pre- and post-test before and after the training program, respectively. The comparison group comprised four cohorts (cohort 1: $n = 11$; cohort 2: $n = 7$; cohort 3: $n = 8$; cohort 4: $n = 13$) resulting in a total of $N = 39$ biology pre-service teachers as participants (82% female). Their mean age was 23.2 ($SD = 2.6$). On average, the participants were in the sixth semester of their university teacher educational program ($M = 6.3$; $SD = 1.8$). They completed the pre- and post-test before and after, respectively, a biology educational seminar or the university-based theory practice term with an interim period of approximately three months. All pre-service teachers were enrolled in a university program at our university to become teachers for secondary schools and participated on a voluntary basis. The requirements of passing the university courses were independent of the participation in this study. All participants received the identical digital introduction to the study and assessment. Due to time economic factors, no data regarding motivational aspects were collected.
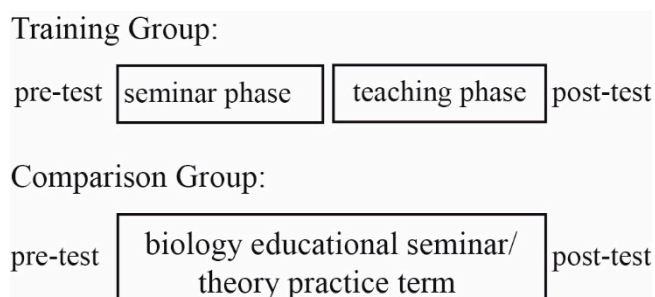

Figure 3. Study design. The training group participated in the new training. The participants of comparison group completed either a biology educational seminar or the theory practice term.

2.4. Analyses of Data

We used item response theory (IRT) models to scale our data. IRT models are often used for data analysis in the field of empirical studies focused on performance tests [23,100–102]. For the analyses of our data, models from the Rasch tradition were used [103,104]. We used R Studio (version 1.0.153) including the TAM package [105] to analyze our data as well as IBM SPSS Statistics (version 24) for further analyses (for details see below).

3. Results

3.1. Assessing Preconditions for Using the Test to Assess Professional Vision

To assess the structure of the data originating from the test, more dimensional models were tested because professional vision seems to consist of four separate processes, namely perception as well as the three processes of interpretation: description, explanation, and prediction (RQ 1). We tested the structure of professional vision by comparing a four-dimensional model presuming that perception, description, explanation, and prediction can be measured as distinct dimensions and more restricted models (Table 2). The more restricted models were a one-dimensional model (pooling all items on one dimension), a two-dimensional model (differentiating the two processes of perception and interpretation),

and a three-dimensional model (distinguishing perception, integrating description/explanation, and prediction). Thus, four different models were tested by contrasting the global model fit of the four-dimensional model with the global model fit of the more restricted models based on the criterion-referenced norms (dichotomous and partial credit).

Table 2. Comparison of different dimensional models for (a) strict and (b) less strict expert reference norm.

	Dimensionality	Deviance	Parameters	Δ Deviance	AIC	BIC
Pre-test	1-D model	12,661	109	162 **	12,879	13,182
	2-D model	12,604	111	104 **	12,826	13,134
	3-D model	12,597	114	97 **	12,825	13,142
	4-D model	12,499	118		12,735	13,063
Post-test	1-D model	13,612	109	259 **	13,830	14,133
	2-D model	13,546	111	193 **	13,768	14,076
	3-D model	13,533	114	180 **	13,761	14,077
	4-D model	13,353	118		13,589	13,917

(a) Strict expert reference norm.

	Dimensionality	Deviance	Parameters	Δ Deviance	AIC	BIC
Pre-test	1-D model	23,021	205	119 **	23,431	24,001
	2-D model	22,958	207	55 **	23,372	23,947
	3-D model	22,945	210	43 **	23,365	23,949
	4-D model	22,902	214		23,330	23,925
Post-test	1-D model	23,377	205	171 **	23,787	24,357
	2-D model	23,312	207	105 **	23,726	24,301
	3-D model	23,300	210	93 **	23,720	24,303
	4-D model	23,207	214		23,635	24,230

(b) Less strict expert reference norm.

Deviance: −2log (likelihood ratio); Parameters: number of free parameters; Δ Deviance: difference in deviance and χ2 distributed test value of the likelihood ratio test. AIC (Akaike's information criterion) and BIC (Bayesian information criterion) are used to estimate the model fit (penalty score) [106] ** $p < 0.01$. For details see text.

The lower penalty scores AIC and BIC for the four-dimensional models indicate that the professional vision assessed with our instrument can be described best with the four-dimensional models (strict and less strict expert-referenced norms) as it fits the data significantly better in relation to the more restricted models. The likelihood ratio test affirmed these results.

The four-dimensional models (strict and less strict expert-referenced norms) were used as a basis for the following investigations. To determine the change in participants' abilities in the repeated measurement design, the procedure of using virtual persons was deployed [107]. Items were analyzed according to the mean square fit index ($0.75 \leq$ MNSQ ≤ 1.30, [108]). As a result, the item pool was reduced from the original $i = 108$ to $i = 92$ items that fit both strict and less strict norm four-dimensional models. Regarding the eliminated items, no meaningful pattern was identifiable, and all relevant aspects of the test were still covered by the remaining items. Hence, we assume that the final test is adequate to assess the professional vision, since no items were eliminated in the dimensions perception ($i = 12$) and explanation ($i = 21$). Due to poor fit, 10 items were eliminated in the dimension explanation resulting in $i = 41$, and six items were eliminated in the dimension prediction resulting in ($i = 18$).

3.2. Test Scoring

To determine which of the two expert-referenced norms is more suitable, the indices of the four-dimensional models were compared (RQ 2). The strict expert-referenced norm (dichotomous model) resulted in better indices, with excellent to good reliabilities for the scales description (EAP = 0.91) and prediction (EAP = 0.85) as well as a good item discrimination with

up to $\sigma^2 = 2.13$ explained variance (Table 3). The less strict expert-referenced norm (partial credit model) showed reliability scores similar to the model with the strict expert-referenced norm but a low discrimination of the scales (explained variance). Due to the unacceptable and poor reliability of the scales for perception and explanation for both norms, these scales need further improvement in the future. Hence, no further results regarding these scales are reported here, and all interpretations and conclusions are limited to the scales description and prediction.

Table 3. Comparison of reliability and variance of the four-dimensional models.

	Strict Expert-Referenced Norm (1 = Hit; 0 = Miss)	Less Strict Expert-Referenced Norm (2 = Hit; 1 = Close; 0 = Miss)
Abilities	**EAP/WLE Reliability**	
Perception	0.36/0.24	0.35/0.24
Description	0.91/0.85	0.91/0.89
Explanation	0.54/0.15	0.60/0.27
Prediction	0.85/0.64	0.80/0.67
	Variance	
Perception	0.20	0.20
Description	2.13	0.49
Explanation	0.11	0.08
Prediction	1.88	0.24

EAP: expected a posteriori reliability; WLE: Warm's likelihood estimates reliability [109]. Both scores are based on the personestimated ability value and represent predictive reliability that can be interpreted much like Cronbach's α [102,110].

3.3. Demonstrating Sensitivity

In order to test the sensitivity of the test instrument (RQ 3), repeated measure ANOVA was conducted, and the pairwise comparison tests were Bonferroni corrected (Figure 4). Regarding the ability of description, we found a significant interaction effect of group (training vs. comparison group) and time ($F(1,117) = 29.14\ p < 0.001$) with a large effect ($\eta^2_{part} = 0.20$). The pairwise comparison test indicated no significant ($p = 0.602$) difference between the two groups in the pre-test; however, there was a significant ($p < 0.001$) increase in the ability of description of the training group in contrast to the comparison group.
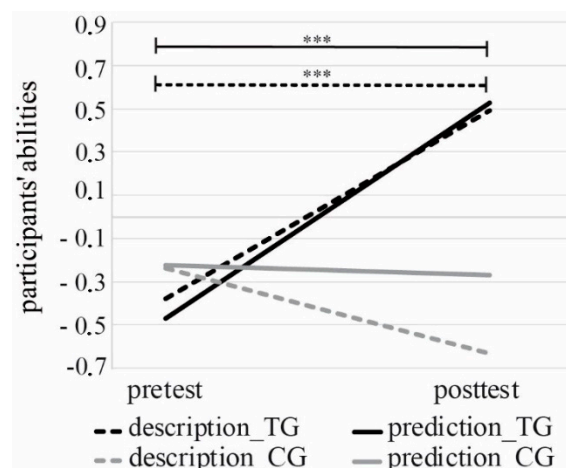

Figure 4. Changes in the abilities describing and predicting of the training group (TG) and the comparison group (CG). *** indicates highly significant ($p < 0.001$) changes.

For the ability of prediction, a significant interaction effect was detected between groups and time ($F(1,117) = 14.81\ p < 0.001$) with a medium effect ($\eta^2_{part} = 0.11$). The pairwise comparison test

revealed that there was no difference (p = 0.377) in the participants' ability of prediction in the pre-test. However, in the post-test the participants' ability to predict was significantly (p = 0.003) higher in the training group in contrast to the comparison group.

4. Discussion

The aim of the present study was to assess biology pre-service teachers' professional vision of formative assessment in the context of scientific inquiry. Adhering to the situated assessment approach of professional vision, authentic text vignettes served as a stimulus. Text vignettes can serve as appropriate stimuli in subject-educational research in particular, as these in contrast to video vignettes are primarily less complex in terms of simultaneously occurring processes, whereas video vignettes are mainly of interest in pedagogical or general educational research [111]. Additionally, Friesen et al. [112] were able to show that the format of the vignettes (video, text, or comic) did not affect the teachers' perception in subject-didactical contexts. Text and video vignettes are both perceived as authentic representations of classroom settings [113]. The authenticity of the vignettes is crucial for the initiation of the professional vision [114]. However, staged videos can also be perceived as authentic [115].

The development of our test instrument was successful in the sense that a four-dimensional construct of professional vision was detected, which is in line with the already described abilities of professional vision (perception, description, explanation, prediction) [25–27]. This finding is an indication of construct validity; however, further support for validity is needed.

Regarding the expert-referenced norm comparison (strict vs. less strict), the strict model is to be favored over the less strict model, since the explained variance was higher in the strict model. In contrast to the scales description and prediction, the scales testing the abilities perception and explanation did not have satisfactory reliability. The low reliability for the ability perception may be due to the low variance caused by the dichotomous item format. Due to content-based considerations, we preferred to use a dichotomous answer format over a four-point Likert scale format as participants either perceive or do not perceive a certain aspect in the vignettes. However, it seems more promising to use a four-point Likert scale format [33] or assessing perception indirectly, since perceiving certain aspects forms the basis for interpretation [23]. The reliability for the ability explanation may have been unsatisfactory based on the low variance present in the participants' responses, which was possibly due to most of the items being too difficult in relation to the participants' abilities. Therefore, the items for perception and explanation have to be revised. In our training we considered the features of successful training regarding professional vision (such as analyzing teaching videos, role plays, and active teaching in microteaching situations) and incorporated the important aspect of theory–practice integration. Hence, we assumed that professional vision of the pre-service teachers could be strengthened [34,35,48,116]. By measuring significant changes in professional vision regarding the abilities description and prediction, the conceptualized instrument is shown to be sensitive enough to detect changes in these abilities of participants' professional vision.

5. Limitations and Outlook

The support for the validity of the instrument is not comprehensive. Further investigations regarding validity evidence could be realized in the future by conducting, for example, cross validation of the results based on a replication study. Concerning the sensitivity of the instrument, verification of the increase of professional vision of pre- and in-service teachers is needed, similar to Gold and Holodynski [91]. Sources of validity evidence based on response processes presenting insights of test-taking population regarding their performance strategies such as eye tracking or the thinking aloud method could help to understand the fit between the construct of professional vision and the actual response the test takers are engaged in. Only two of four scales showed satisfying reliability, as the scales for the abilities perception and explanation need further refinements. Thus, a holistic assessment of professional vision by means of the instrument is not possible right now. Furthermore, the content

specificity of the instrument has to be taken into account in future studies; hence, generic statements regarding professional vision are not suitable based on the data collected by the present instrument.

6. Conclusions

This paper reports the investigation of a text-vignette-based test instrument focusing on professional vision of formative assessment regarding scientific inquiry. Data assessed with this test instrument fit the theoretically assumed four-dimensional structure of professional vision (RQ 1). We found that the use of a strict expert-referenced norm is to be favored over the partial credit model (RQ 2). Furthermore, the test instrument was sensitive enough to detect changes in the professional vision of pre-service teachers that participated in training sessions focused on professional vision of formative assessment regarding scientific inquiry concerning the abilities description and prediction (RQ 3). Incorporating the instrument in existing out-of-school laboratory courses will improve understanding of their effects on the development of professional vision, thereby enabling insightful comparisons with other kinds of courses.

References

1. Shulman, L.S. Knowledge and teaching: Foundations of new reform. *Harv. Educ. Rev.* **1987**, *57*, 1–21. [CrossRef]
2. Großschedl, J.; Mahler, D.; Kleickmann, T.; Harms, U. Content-related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *Int. J. Sci. Educ.* **2014**, *36*, 2335–2366. [CrossRef]
3. Jüttner, M.; Neuhaus, B.J. Development of items for a pedagogical content knowledge test based on empirical analysis of pupils' errors. *Int. J. Sci. Educ.* **2012**, *34*, 1125–1143. [CrossRef]
4. Käpylä, M.; Heikkinen, J.-P.; Asunta, T. Influence of content knowledge on pedagogical content knowledge: The case of teaching photosynthesis and plant growth. *Int. J. Sci. Educ.* **2009**, *31*, 1395–1415. [CrossRef]
5. Park, S.; Chen, Y.-C. Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *J. Res. Sci. Teach.* **2012**, *49*, 922–941. [CrossRef]
6. Rozenszajn, R.; Yarden, A. Expansion of biology teachers' pedagogical content knowledge (PCK) during a long-term professional development program. *Res. Sci. Educ.* **2014**, *44*, 189–213. [CrossRef]
7. Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M.; Tsai, Y.-M. Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *Am. Educ. Res. J.* **2010**, *47*, 133–180. [CrossRef]
8. Hill, H.C.; Rowan, B.; Ball, D.L. Effects of teachers' mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* **2005**, *42*, 371–406. [CrossRef]
9. Hashweh, M.Z. Effects of science teachers' epistemological beliefs in teaching. *J. Res. Sci. Teach.* **1996**, *33*, 47–63. [CrossRef]
10. Lord, T.R. A comparison between traditional and constructivist teaching in college biology. *Innov. High. Educ.* **1997**, *21*, 197–216. [CrossRef]
11. Luft, J.A. Changing inquiry practices and beliefs: The impact of an inquiry-based professional development programme on beginning and experienced secondary science teachers. *Int. J. Sci. Educ.* **2001**, *23*, 517–534. [CrossRef]

12. Luft, J.A.; Hewson, P.W. Research on teacher professional development programs in science. In *Handbook of Research on Science Education*; Ledermann, N.G., Abell, S.K., Eds.; Routledge: New York, NY, USA, 2014; pp. 889–909.

13. Luft, J.A.; Roehring, G.H. Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. *Electron. J. Sci. Educ.* **2007**, *11*, 38–63.

14. Seidel, T.; Schwindt, K.; Rimmele, R.; Prenzel, M. Konstruktivistische Überzeugungen von Lehrpersonen: Was bedeuten sie für den Unterricht? Constructivist beliefs of teachers: Which impact do they have on school lessons? In *Perspektiven der Didaktik: Zeitschrift für Erziehungswissenschaft*; Meyer, M.A., Hellekamps, S., Prenzel, M., Eds.; VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH: Wiesbaden, Germany, 2009; pp. 259–276.

15. Gess-Newsome, J.; Lederman, N.G. Biology Teachers' Perceptions of Subject Matter Structure and its Relationship to Classroom Practice. *J. Res. Sci. Teach.* **1992**, *32*, 301–325. [CrossRef]

16. Mellado, V. Preservice teachers' classroom practice and their conceptions of the nature of science. *Sci. Educ.* **1997**, *6*, 331–354. [CrossRef]

17. Treagust, D.F.; Duit, R.; Joslin, P.; Lindauer, I. Science teachers' use of analogies: Observations from classroom practice. *Int. J. Sci. Educ.* **1992**, *14*, 413–422. [CrossRef]

18. Wenglinsky, H. How Schools Matter: The link between teacher classroom practices and student academic performance. *Educ. Policy Anal. Arch.* **2002**, *10*, 1–30.

19. Blömeke, S.; Gustafsson, J.-E.; Shavelson, R.J. Beyond dichotomies: Competence viewed as a continuum. *Z. Für Psychol.* **2015**, *223*, 3–13.

20. Seidel, T.; Blomberg, G.; Stürmer, K. "Observer"—Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. Schwerpunktprogramms und Perspektiven des Forschungsansatzes ["Observer"—Validation of a video-based instrument for capturing the professional vision of school lessons. Projekt OBSERVE]. In *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik*; Klieme, E., Leutner, D., Kenk, M., Eds.; Beltz: Weinheim, Germany, 2010; pp. 296–306.

21. Sherin, M. The development of teachers' professional vision in teacher clubs. In *Video Research in the Learning Sciences*; Goldman, R., Ed.; Erlbaum: Mahwah, NJ, USA, 2007; pp. 383–395.

22. van Es, E.A.; Sherin, M.G. Mathematics teachers' "learning to notice" in the context of a video club. *Teach Teach. Educ.* **2008**, *24*, 244–276. [CrossRef]

23. Seidel, T.; Stürmer, K. Modeling and measuring the structure of professional vision in preservice teachers. *Am. Educ. Res. J.* **2014**, *51*, 739–771. [CrossRef]

24. van Es, E.A.; Sherin, M.G. Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *J. Technol. Teach. Educ.* **2002**, *10*, 571–596.

25. Sherin, M.G.; van Es, E.A. Effects of Video Club Participation on Teachers' Professional Vision. *J. Teach. Educ.* **2009**, *60*, 20–37. [CrossRef]

26. Berliner, D.C. Learning about and learning from expert teachers. *Int. J. Educ. Res.* **2001**, *35*, 463–482. [CrossRef]

27. Borko, H.; Livingston, C. Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *Am. Educ. Res. J.* **1989**, *26*, 473–498. [CrossRef]

28. Schwindt, K. *Teachers Observe Instruction: Criteria for Competent Perception of Instruction*; Waxmann: Münster, Germany, 2008.

29. Zumbach, J.; Haider, K.; Mandl, H. Fallbasiertes Lernen: Theoretischer Hintergrund und praktische Anwendung [Case-based learning: Theoretical background and practical application]. In *Pädagogische Psychologie in Theorie und Praxis: Ein Fallbasiertes Lehrbuch*; Mandl, H., Zumbach, J., Eds.; Hogrefe: Göttingen, Germany; Bern, Switzerland, 2008; pp. 1–11.

30. Borko, H.; Livingston, C.; Shavelson, R.J. Teachers' thinking about instruction. *Remedial Spec. Educ.* **1990**, *11*, 40–49. [CrossRef]

31. Stürmer, K.; Könings, K.D.; Seidel, T. Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *Br. J. Educ. Psychol.* **2013**, *83*, 467–483. [CrossRef] [PubMed]

32. Goodwin, C. Professional vision. *Am. Anthropol.* **1994**, *96*, 606–633. [CrossRef]

33. Steffensky, M.; Gold, B.; Holdynski, M.; Möller, K. Professional vision of classroom management and learning support in science classrooms—Does professional vision differ across general and content-specific classroom interactions? *Int. J. Sci. Math. Educ.* **2015**, *13*, 351–368. [CrossRef]

34. Star, J.R.; Strickland, S.K. Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *J. Math. Teach. Educ.* **2008**, *11*, 107–125. [CrossRef]

35. Sunder, C.; Todorova, M.; Möller, K. Kann die professionelle Unterrichtswahrnehmung von Sachunterrichtsstudierenden trainiert werden?—Konzeption und Erprobung einer Intervention mit Videos aus dem naturwissenschaftlichen Grundschulunterricht [Can social studies students' professional vision be trained? Concepting and testing an intervention with videos from scientific primary school education]. *Z. Für Didakt. Der Nat.* **2016**, *22*, 1–12.

36. Beck, R.J.; King, A.; Marshall, S.K. Effects of videocase construction on preservice teachers' observations of teaching. *J. Exp. Educ.* **2002**, *4*, 345–361. [CrossRef]

37. Michalsky, T. Developing the SRL-PV assessment scheme: Preservice teachers' professional vision for teaching self-regulated learning. *Stud. Educ. Eval.* **2014**, *43*, 214–229. [CrossRef]

38. Santagata, R.; Zannoni, C.; Stigler, J.W. The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *J. Math. Teach. Educ.* **2007**, *10*, 123–140. [CrossRef]

39. Möller, K.; Steffensky, M.; Meschede, N.; Wolters, M. Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht [Professional vision of learning support within scientific primary school education]. *Unterrichtswissenschaft* **2015**, *43*, 317–335.

40. Wolff, C.E.; Jarodzka, H.; van den Bogert, N.; Boshuizen, H.P.A. Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instr. Sci.* **2016**, *44*, 243–265. [CrossRef]

41. Wolff, C.E.; Jarodzka, H.; Boshuizen, H. See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teach. Teach. Educ.* **2017**, *66*, 295–308. [CrossRef]

42. Stürmer, K.; Seidel, T.; Müller, K.; Häusler, J.S.; Cortina, K. What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Z. Für Erzieh.* **2017**, *20*, 75–92.

43. Dreher, A.; Kuntze, S. Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educ. Stud. Math.* **2015**, *88*, 89–114. [CrossRef]

44. Son, J.-W. How preservice teachers interpret and respond to student errors: Ration and proportion in similar rectangles. *Educ. Stud. Math.* **2013**, *84*, 49–70. [CrossRef]

45. Kersting, N.B.; Givvin, K.B.; Thompson, B.J.; Santagata, R.; Stigler, J.W. Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *Am. Educ. Res. J.* **2012**, *49*, 568–589. [CrossRef]

46. Roth, K.J.; Garnier, H.E.; Chen, C.; Lemmens, M.; Schwille, K.; Wickler, N.I.Z. Videobased lesson analysis: Effective science PD for teacher and student learning. *J. Res. Sci. Teach.* **2011**, *48*, 117–148. [CrossRef]

47. Hamre, B.K.; Pianta, R.C.; Burchinal, M.; Field, S.; LoCasale-Crouch, J.; Downer, J.T.; Howes, C.; LaParo, K.; Scott-Little, C. A Course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *Am. Educ. Res. J.* **2012**, *49*, 88–123. [CrossRef]

48. Gold, B.; Förster, S.; Holodynski, M. Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht [Evaluation of a video-based training seminar for supporting classroom managements' professional vision in primary school education]. *Z. Für Pädagogische Psychol.* **2013**, *27*, 141–155.

49. Santagata, R.; Guarino, J. Using video to teach future teachers to learn from teaching. *Zdm Math. Educ.* **2011**, *43*, 133–145. [CrossRef]

50. Barth, V.L.; Piwowar, V.; Kumschick, I.R.; Ophardt, D.; Thiel, F. The impact of direct instruction in a problem-based learning setting. Effects of a video-based training program to foster preservice teachers' professional vision of critical incidents in the classroom. *Int. J. Educ. Res.* **2019**, *95*, 1–12. [CrossRef]

51. Alsawaie, O.N.; Alghazo, I.M. The effect of video-based approach on prospective teachers' ability to analyze mathematics teaching. *J. Math. Teach. Educ.* **2010**, *13*, 223–241. [CrossRef]

52. Roth, K.J.; Bintz, J.; Wickler, N.I.Z.; Hvidsten, C.; Taylor, J.; Beardsley, P.M.; Caine, A.; Wilson, C.D. Design principles for effective video-based professional development. *Int. J. Stem Educ.* **2017**, *4*, 31. [CrossRef]

53. Tekkumru-Kisa, M.; Stein, M.K. Designing, facilitating, and scaling-up video-based professional development: Supporting complex forms of teaching in science and mathematics. *Int. J. Stem Educ.* **2017**, *4*, 27. [CrossRef]

54. Tekkumru-Kisa, M.; Stein, M.K. A framework for planning and facilitating video-based professional development. *Int. J. Stem Educ.* **2017**, *4*, 28. [CrossRef]

55. Krammer, K.; Ratzka, N.; Klieme, E.; Lipowsky, F.; Pauli, C.; Reusser, K. Learning with classroom videos: Conception and first results of an online teacher-training program. *Zent. Für Didakt. Math.* **2006**, *38*, 422–432. [CrossRef]

56. Blomberg, G.; Renkl, A.; Sherin, M.G.; Borko, H.; Seidel, T. Five research-based heuristics for using video in pre-service teacher education. *J. Educ. Res. Online* **2013**, *5*, 90–114.

57. Gaudin, C.; Chaliès, S. Video viewing in teacher education and professional development: A literature review. *Educ. Res. Rev.* **2015**, *16*, 41–67. [CrossRef]

58. Coffey, A.M. Using Video to develop skills in reflection in teacher education students. *Aust. J. Teach. Educ.* **2014**, *39*, 86–97. [CrossRef]

59. Shernoff, D.J.; Sinha, S.; Bressler, D.M.; Ginsburg, L. Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education. *Int. J. Stem Educ.* **2017**, *4*, 13. [CrossRef] [PubMed]

60. Gold, B.; Pfirrmann, C.; Holodynski, M. Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *J. Teach. Educ.* **2020**, 002248712096368. [CrossRef]

61. Weber, K.E.; Gold, B.; Prilop, C.N.; Kleinknecht, M. Promoting pre-service teachers' professional vision of classroom management during practical school training: Effects of a structured online- and video-based self-reflection and feedback intervention. *Teach. Teach. Educ.* **2018**, *76*, 39–49. [CrossRef]

62. National Research Council. *National Science Education Standards*; National Academy Press: Washington, DC, USA, 1996.

63. National Research Council. *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*; National Academies Press: Washington, DC, USA, 2000.

64. The Australian Curriculum F—10. Available online: https://www.australiancurriculum.edu.au/download/DownloadF10 (accessed on 14 November 2020).

65. Department for Education and Skills/Qualification and Curriculum Authority. *Handbook for Secondary Teachers in England (2004): The National Curriculum: Key Stages $\frac{3}{4}$*; HMSO: London, UK, 2004.

66. Ledermann, N.G.; Abell, S.K. *Handbook of Research on Science Education*; Routledge: New York, NY, USA, 2014.

67. Maier, M. *Entwicklung und Prüfung eines Instrumentes zur Diagnose der Experimentierkompetenz von Schülerinnen und Schülern [An Instruments' Development and Examination for Diagnosing the Experimenting Competence of Pupils]*; Logos Verlag: Berlin, Switzerland, 2015.

68. Germann, P.J.; Aram, R.; Burke, G. Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *J. Res. Sci. Teach.* **1996**, *33*, 79–99. [CrossRef]

69. Hammann, M.; Phan, T.T.H.; Ehmer, M.; Bayhuber, H. Schulpraxis-Fehlerfrei Experimentieren [Experimenting correctly]. *Math. Nat. Unterr.* **2006**, *59*, 292–299.

70. Chen, Z.; Klahr, D. All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Dev.* **1999**, *70*, 1098–1120. [CrossRef]

71. Klahr, D.; Fay, A.L.; Dunbar, K. Heuristics for scientific experimentation: A developmental study. *Cogn. Psychol.* **1993**, *25*, 111–146. [CrossRef]

72. Arnold, J.C.; Kremer, K.; Mayer, J. Understanding students' experiments—What kind of support do they need in inquiry tasks? *Int. J. Sci. Educ.* **2014**, *36*, 2719–2749. [CrossRef]

73. Bybee, R.W. Teaching science as inquiry. In *Inquiry into Inquiry Learning and Teaching in Science*; Minstrell, J., van Zee, E.H., Eds.; American Association for the Advancement of Science: Washington, DC, USA, 2000; pp. 20–46.

74. Furtak, E.M. The problem with answers: An exploration of guided scientific inquiry teaching. *Sci. Ed.* **2006**, *90*, 453–467. [CrossRef]

75. Mayer, R.E. Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *Am. Psychol.* **2004**, *59*, 14–19. [CrossRef] [PubMed]

76. Kirschner, P.A.; Sweller, J.; Clark, R.E. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* **2006**, *41*, 75–86. [CrossRef]

77. Bell, B.; Cowie, B. The characteristics of formative assessment in science education. *Sci. Educ.* **2001**, *85*, 536–553. [CrossRef]

78. Black, P.; Wiliam, D. Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* **2009**, *21*, 5–31. [CrossRef]

79. Decristan, J.; Hondrich, A.L.; Büttner, G.; Hertel, S.; Klieme, E.; Kunter, M.; Lühken, A.; Adl-Amini, K.; Djarkovic, S.; Mannel, S.; et al. Impact of additional guidance in science education on primary students' conceptual understanding. *J. Educ. Res.* **2015**, *108*, 358–370. [CrossRef]

80. Shavelson, R.J.; Young, D.B.; Ayala, C.C.; Brandon, P.R.; Furtak, E.M.; Ruiz-Primo, M.; Tomita, M.K.; Yin, Y. On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Appl. Meas. Educ.* **2008**, *21*, 295–314. [CrossRef]

81. Hattie, J. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*; Routledge: New York, NY, USA, 2008.

82. Decristan, J.; Klieme, E.; Kunter, M.; Hochweber, J.; Büttner, G.; Fauth, B.; Hondrich, A.L.; Rieser, S.; Hertel, S.; Hardy, I. Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *Am. Educ. Res. J.* **2015**, *52*, 1133–1159. [CrossRef]

83. Loughland, T.; Kilpatrick, L. Formative assessment in primary science. *Int. J. Prim. Elem. Early Years Educ.* **2013**, *43*, 128–141. [CrossRef]

84. Ruiz-Primo, M.A.; Furtak, E.M. Informal formative assessment and scientifiv inquiry: Exploring teachers' practices and student learning. *Educ. Assess.* **2006**, *11*, 205–235.

85. Coffey, J.E.; Hammer, D.; Levin, D.M.; Grant, T. The missing disciplinary substance of formative assessment. *J. Res. Sci. Teach.* **2011**, *48*, 1109–1136. [CrossRef]

86. Gotwals, A.W.; Philhower, J.; Cisterna, D.; Bennett, S. Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers. *Int. J. Sci. Math. Educ.* **2015**, *13*, 405–423. [CrossRef]

87. Morrison, J.A.; Lederman, N.G. Science teachers' diagnosis and understanding of students' preconceptions. *Sci. Educ.* **2003**, *87*, 849–867. [CrossRef]

88. Davis, E.A.; Petish, D.; Smithey, J. Challenges new science teachers face. *Rev. Educ. Res.* **2006**, *76*, 607–651. [CrossRef]

89. Furtak, E.M.; Ruiz-Primo, M.A.; Bakeman, R. Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educ. Meas. Issues Pract.* **2017**, *36*, 28–38. [CrossRef]

90. Allen, D.W.; Eve, A.W. Microteaching. *Theory Pract.* **1968**, *7*, 181–185. [CrossRef]

91. Gold, B.; Holodynski, M. Using digital video to measure the professional vision of elementary classroom management: Test validation and methodological challenges. *Comput. Educ.* **2017**, *107*, 13–30. [CrossRef]

92. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [CrossRef]

93. Wirtz, M.A.; Caspar, F. *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen [Rater Agreement and Rater Reliability: Methods for Identification and Improvement of Ratings Using Category Systems and Rating Scales]*; Hogrefe Verl. für Psychologie: Göttingen, Germany, 2002.

94. Baumann, S. *Selbständiges Experimentieren und Konzeptuelles Lernen mit Beispielaufgaben in Biologie [Self-Reliant Experimentation and Conceptual Learning by Using Worked Examples in Biology]*; Logos Verlag: Berlin, Germany, 2014.

95. Haugwitz, M.; Sandmann, A. Collaborative modelling of the vascular system - designing and evaluating a new learning method for secondary students. *J. Biol. Educ.* **2010**, *44*, 136–140. [CrossRef]

96. Rumann, S. *Kooperatives Experimentieren im Chemieunterricht: Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik*; Logos: Berlin, Germany, 2005.

97. Vogt, F.; Schmiemann, P. Development of professional vision, knowledge, and beliefs of pre-service teachers in an out-of-school laboratory course. In *Professionalisierung durch Lehr-Lern-Labore in der Lehrerausbildung*; Bosse, D., Meier, M., Trefzger, T., Ziepprecht, K., Eds.; Verlag Empirische Pädagogik: Landau in der Pfalz, Germany, 2020; pp. 25–47.

98. Mayer, J. Erkenntnisgewinnung als wissenschaftliches Problemlösen [Knowledge acquisition as scientific problem solving]. In *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden*; Krüger, D., Vogt, H., Eds.; Springer-Lehrbuch: Berlin/Heidelberg, Germany, 2007; pp. 177–196.

99. Zhang, M.; Lundeberg, M.; Koehler, M.J.; Eberhardt, J. Understanding affordances and challenges of three types of video for teacher professional development. *Teach. Teach. Educ.* **2011**, *27*, 454–462. [CrossRef]

100. Boone, W.J.; Scantlebury, K. The role of Rasch Analysis When Conducting Science Education Research Utilizing Multiple-Choice Tests. *Sci. Ed.* **2006**, *90*, 253–269. [CrossRef]

101. Harrison, G.M.; Duncan Seraphin, K.; Philippoff, J.; Vallin, L.M.; Brandon, P.R. Comparing Models of Nature of Science Dimensionality Based on the Next Generation Science Standards. *Int. J. Sci. Educ.* **2015**, *37*, 1321–1342. [CrossRef]

102. Neumann, I.; Neumann, K.; Nehm, R. Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *Int. J. Sci. Educ.* **2011**, *33*, 1373–1405. [CrossRef]

103. Rasch, G. *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*; Nielsen & Lydiche: Oxford, UK, 1960.

104. Masters, G.N. A rasch model for partial credit scoring. *Psychometrika* **1982**, *47*, 149–174. [CrossRef]

105. Kiefer, T.; Robitzsch, A.; Wu, M.; Robitzsch, M.A. Package 'TAM'. Available online: https://mran.microsoft.com/snapshot/2017-02-20/web/packages/TAM/TAM.pdf (accessed on 16 June 2020).

106. Burnham, K.P.; Anderson, D.R. Multimodel Inference. *Sociol. Methods Res.* **2004**, *33*, 261–304. [CrossRef]

107. Hartig, J.; Kühnbach, O. Schätzung von Veränderung mit "plausible values" in mehrdimensionalen Rasch-Modellen [Estimation of changes with "plausible values" within the multidimensional Rasch-models]. In *Veränderungsmessung und Längsschnittstudien in der Empirischen Erziehungswissenschaft*, 1st ed.; Ittel, A., Ed.; VS Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2006; pp. 27–44.

108. Bond, T.G.; Fox, C.M. *Applying the Rasch Model*; Erlbaum: Mahwah, NJ, USA, 2001.

109. Warm, T.A. Weighted likelihood estimation of ability in item response theory. *Psychometrika* **1989**, *54*, 427–450. [CrossRef]

110. Kim, J.K.; Nicewander, W.A. Ability estimation for conventional test. *Pyschometrika* **1993**, *58*, 587–599. [CrossRef]

111. Brovelli, D.; Bölsterli, K.; Rehm, M.; Wilhelm, M. Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettentest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat [Recording professional competencies for science classes: A vignette-test with authentic and complex teaching situations and open response format]. *Unterrichtswissenschaft* **2013**, *41*, 306–329.

112. Friesen, M.; Kuntze, S.; Vogel, M. Videos, Texte oder Comics? Die Rolle des Vignettenformats bei der Erhebung fachdidaktischer Analysekompetenz zum Umgang mit Darstellungen im Mathematikunterricht [Videos, texts or comics? The role of the vignette-format when elevating didactical analysis competences dealing with representations in mathematics classes]. In *Effektive Kompetenzdiagnose in der Lehrerbildung: Professionalisierungsprozesse Angehender Lehrkräfte Untersuchen*; Rutsch, J., Rehm, M., Vogel, M., Seidenfuß, M., Dörfler, T., Eds.; Springer: Wiesbaden, Germany, 2018; pp. 153–177.

113. Zucker, V. *Erkennen und Beschreiben von formativem Assessment im Naturwissenschaftlichen Grundschulunterricht: Entwicklung eines Instruments zur Erfassung von Teilfähigkeiten der professionellen Wahrnehmung von Lehramtsstudierenden [Perception and Description of Formative Assessment in Scientific Primary School Education: Development of an Instrument to Assess Sub-Abilities of Professional Vision of Pre-Service Teachers]*; Logos Verlag: Berlin, Germany, 2019.

114. Stürmer, K.; Seidel, T. Connecting generic pedagogical knowledge with practice. In *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*; Guerriero, S., Ed.; OECD Publishing: France, Paris, 2017; pp. 137–149.

115. Kramer, M.; Förtsch, C.; Stürmer, J.; Förtsch, S.; Seidel, T.; Neuhaus, B.J. Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Educ.* **2020**, *7*, 1823155. [CrossRef]

116. Stürmer, K.; Seidel, T.; Holzberger, D. Intra-individual differences in developing professional vision: Preservice teachers' changes in the course of an innovative teacher education program. *Instr. Sci.* **2016**, *44*, 293–309. [CrossRef]