© Universiti Tun Hussein Onn Malaysia Publisher's Office

JTET

http://penerbit.uthm.edu.my/ojs/index.php/jtet
ISSN 2229-8932 e-ISSN 2600-7932

Journal of Technical Education and Training

Determination of a Reference Value for Adequate Assessment of Teaching Situations: Development of a Technical Education Expert Norm (PCK-T)

Sebastian Goreth[1*], Friederike Straub[2], Markus Rehm[3], Bernd Geißel[4]

[1]University of Education Tirol, Pastorstraße 7, Innsbruck, 6020, AUSTRIA

[2,4]University of Education Ludwigsburg, Reuteallee 46, Ludwigsburg, 71634, GERMANY

[3]University of Education Heidelberg, Keplerstraße 87, Heidelberg, 69120, GERMANY

*Corresponding Author

DOI: https://doi.org/10.30880/jtet.2020.12.04.008
Received 15th April 2020; Accepted 10th August 2020; Available online 31st December 2020

Abstract: The different facets of professional performance of teachers are being debated in the current educational research discussion. Both qualitative and quantitative test constructions exist for vignette-based measurement of competence. For economic reasons, a procedure with closed-ended items is often favored, a reference is required for determining and comparing alternative responses. This paper sketches out the procedure for identifying adequate reactions to responses that we call a technical education expert norm (PCK-T). This expert norm is generated from a multi-step expert survey and, in addition, reveals possible validation steps that can be derived for developing teaching situations. After a content validation ($N_1 = 8$) a multi-step quantitative survey with specialised subject experts at schools and universities, departmental heads at public colleges for education and teacher training, as well as experienced teachers of technology was carried out ($N_2 = 79$; $N_3 = 76$). In order to assess teaching competencies of pre-service teachers using a vignette-based test procedure, the generated technical education expert norm (PCK-T) allows adequate responses to be differentiated from (rather) inadequate responses in the teaching situations.

Keywords: Vignettes, technical didactic expertise, PCK, technical education expert norm

1. Introduction

Enabling students to achieve technical maturity – i.e., the ability to use technological developments, as well as to assess and evaluate their consequences for themselves, society and the environment (VDI, 2012) – represents a complex demand of teaching practice for teachers of technical education. Achieving maturity implies the assumption of responsibility, as well as appropriate action in a world that is shaped by technology. To this end, a high degree of professionalism in the classroom is required of teachers. In Germany, however, many different conceptualizations of technical instruction are to be found in classroom teaching. In addition to subject groupings like *Science and Technology*, *Nature and Science*, and *Economics and Science*, the subject *Technology* also exists as a discipline in some of the German *federal states* and is thus marked by shifting political expectations or contents (Tenberg, 2016; Zinn, 2014). Moreover, the different disciplinary cultures have to be distinguished with respect to both school form and age. This has as consequence that teacher training takes different shapes in the different states. What appears to be common to all of them is that technical instruction is distinguished from other school subjects by its concrete orientation to practice (Traebert, 2001).

*Corresponding author: sebastian.goreth@ph-tirol.ac.at

In countries like Switzerland or, for instance, in Norway and England, integrated forms of studies of science education can be found for teachers for example natural sciences or “general science” (Rehm et al., 2008). In Germany, specific disciplinary training in studies of natural sciences is largely preferred. Although some aspects argue in favour of an integrated approach, for the time being empirical evidence on the effectiveness of the different training structures is lacking (Tenberg, 2014). The present research project takes up this gap in research and compares pedagogical content competencies in different subject combinations. To this end, this paper sketches out the possibility of generating a reference value (Oser, 2015) in the form of a technical education expert norm (PCK-T), in order to assess teaching competencies of pre-service teachers using a vignette-based test procedure. In addition, derived validation steps in the process of test development are presented. In the further course of the project, student answers will be placed in relation to the generated expert norm, assigned points, and evaluated (Goreth, 2017). This will not, however, be the subject of the present paper.

2. Theoretical Background

According to Shulman (1987), the teaching profession begins “with a teacher’s understanding of what is to be learned and how it is to be taught” (Shulman, 1987, p. 7). Apart from the understanding of the teaching content, as well as the sort of learning, in his schema of classification, he also distinguishes seven areas within the professional knowledge of teachers: content knowledge, general pedagogical knowledge, curriculum knowledge, pedagogical content knowledge, knowledge of learners, knowledge of educational contexts, and knowledge of educational ends (Shulman, 1987). In the model of Baumert and Kunter (2013), as well as Voss, Kunina-Habenicht, Hoehne and Kunter (2015), three facets of professional knowledge are ascribed a higher level of significance: content knowledge (CK), general pedagogical knowledge ((G)PK), and pedagogical content knowledge (PCK) (Kaiser, Busse, Hoth, König & Blömeke, 2015; König, 2015; Abell, 2007).

In day-to-day classroom situations, teachers often face more complex demands, which cannot simply be traced back to drawing on different stocks of knowledge. *A professional vision of teaching* appears to be the prerequisite for adequate teaching action. In new models, this expands the teaching process of instructional staff, taking into account experiences and conceptions of teaching and learning (Meschede, Steffensky, Wolters & Möller, 2015; Steffensky, Gold, Holdynski & Möller, 2015; Stürmer & Seidel, 2015; Meschede, 2014). The assumption is that experienced teachers can perceive the process of teaching as a whole, while, at the same time, placing the focus on particular areas, whereas novices are only able to focus on particular areas (Figure 1).

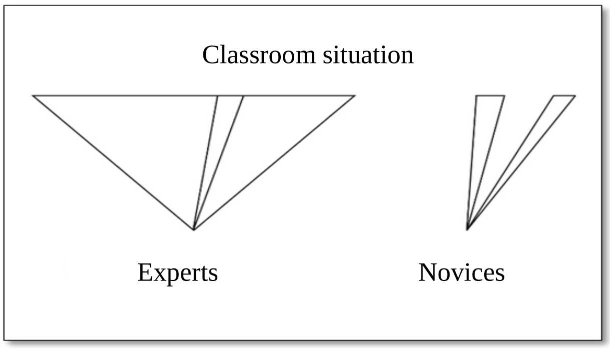

Fig. 1 - Perception of the teaching process from the point of view of experts and novices (Goreth et al., 2016, p. 37).

In recent years, research studies of the competent instructional activity of teachers have been the focus (König, 2015). Thus, a few studies already exist for the domain of mathematics teaching, which, for example, investigate practical aspects of professional knowledge (e.g., Bruckmaier & Krauss, 2015; Kunter, Baumert, Blum, Klusmann, Krauss & Neubrand, 2013; Hugener, Reusser, Lipowsky, Rakoczy & Klieme, 2009). Within the teaching of natural sciences as well, some starting points are already available from projects dealing with the identification of knowledge aspects, but also of professional teaching competencies (*ProwiN*, *ProfiLE-P* and *Ko-Wadis*; Riese et al., 2015; Tepner and Dollny, 2014). Up to now, empirical findings or test constructions on this subject within technology-related teaching exist almost exclusively for the domain of primary school teaching (Rohaan, Taconis & Jochems, 2012; Rohaan, 2009).

In order to develop a test procedure for assessing the subject-specific teaching competencies of pre-service teachers for technical education, a multi-step expert survey design was used. According to the classification of Atteslander (2010), expert surveys are situated as a semi-structured 4th type of survey per the criteria of *degree of communication sort and structuring* (Goreth, Schray, Rehm & Geißel, 2016; Atteslander, 2010) and they refer to persons with a particular expertise.

According to Posner (1988), an expert is a person who persistently (hence not randomly and in a single instance) achieves outstanding performance in a given area (Posner, 1988). Successful mastery of professional demands is characterized as an extensive and inter-connected domain-specific knowledge basis, which can also be called expertise (Baumert & Kunter, 2011). Gruber and Mandl (1996) identify the following six criteria of expertise:

i) large knowledge basis;
ii) a wealth of experience with domain-specific demands;
iii) above-average success in recognizing and dealing with problems;
iv) meta-cognitive monitoring of actions;
v) efficiency, accuracy, and precision of actions;
vi) high flexibility for new problem situations (Gruber & Mandl, 1996).

Accordingly, experts are able to deploy their expertise quickly, effectively and in a solution-oriented fashion in the intended content domain, to think about problems (Hasselhorn & Gold, 2013). Particularly as compared to novices, an expert usually has positive connotations. But this position of superiority only applies within the domains in which the expertise is present (Gruber & Mandl, 1996). Consequently, experts and novices have differing levels of professionalism (Rehm & Bölsterli, 2014). Frey and Jung (2011) regard, above all, the more efficient working methods of experts, as opposed to novices, as the central aspect here. Following the studies in cognitive psychology of Sternberg and Horvarth (1995), these are dependent upon situational schemata that are acquired by way of routines and automatisms in professional practice (Stender, 2014; Frey & Jung, 2011). Hasselhorn and Gold (2013) add that domain-specific expertise is the result of cumulative learning processes, which can be characterised by targeted and guided practice (Hasselhorn & Gold, 2013). The content knowledge must thus be embedded in comprehensive contexts of experience. Hence, extensive content knowledge alone does not lead to a high degree of expertise. The expert must also have had appropriate learning experiences, which, in combination with content knowledge from experiential knowledge, can be readily translated into action (Gruber & Mandl, 1996). This reflects the fact that someone's domain-specific expertise is less an innate competence than a learned one (Hasselhorn & Gold, 2013).

Current research groups in teaching competence research bring together experts who usually have many years of professional experience, since context-specific expertise is assumed precisely in the domain of teaching and schooling. Thus, people can be drawn upon who are active in the working fields of teacher training: such as specialists in education from schools, schools of education and universities, departmental directors and departmental advisors from teaching training and development colleges, and representatives of the respective educational authorities and experienced teachers.

3. Development of a Subject-Specific Didactic Expert Norm

3.1. Competence Assessment Using Vignette-Based Test Procedures

Vignette-based[1]tests have thus far proven to be the most appropriate method for assessing teaching competencies. Apart from distinguishing themselves from limited self-assessment formats, such procedures aim to identify competencies in an *objective* and *proximal* way (Goreth, Geißel & Rehm, 2015; Kunter & Klusmann, 2010). Some research groups use this medium, among other things, incorporating video-based vignette elements (for example, Kaiser, Busse, Hoth, König & Blömeke, 2015; Blomberg, Sherin, Renkl, Glogger & Seidel, 2014; Rehm & Bölsterli, 2014; Riegel, 2013). Moreover, findings on competence processes could already be shown in the course of the study (for example, Brovelli, Bölsterli, Rehm & Wilhelm, 2013; Voss, Kunter & Baumert, 2011).

In the present technical education research project "Assessment and Modeling of Subject-Specific Teaching Competencies in Natural Science and Technology (*EKoL 7*)," a text-based and video-based vignette format has been developed (Goreth, Geißel & Rehm, 2015). The vignettes generated serve to increase the focus on what are regarded as key aspects of technology-related teaching – 1. *Lesson Structure* and 2. *Handling of Tools and Machines* (Goreth, Geißel & Rehm, 2015; Goreth, 2015; VDI, 2007) – and they include closed-ended response-items on a six-point rating scale (1 = "strongly disagree" to 6 = "strongly agree") (Wirtz, 2020). Drawing on the procedure of past projects, an expert norm is determined to establish the state of the students' competence, in order to be able to consider the competences comparatively and thus rate them (Tepner & Dollny, 2014; Tepner et al., 2012; Witner & Tepner, 2011; Seidel, Blomberg and Stürmer, 2010).

There are different approaches to identifying competence. Whereas expert surveys usually only take place once and are not subject to any differentiated methodology (Häder, 2009), an *Expert Delphi* can be characterised as a highly systematic group communication process. As a rule, it is distinguished by several survey steps, by the submission of feedback on the previous round, and by the cognitive activation of the respondents – all of which takes place in a highly standardised survey setting (Häder, 2015; Häder, 2009). The main objective is improved problem-solving. Thus, there are differing variants, which each focus on a different object of investigation in five steps (Häder, 2015; Lehmann, 2012; Beutel, Gröschner & Lütgert, 2006); however, all variants have in common uncertain matters that with the help of this method are meant to be brought closer to an explanation (Häder, 2015).

[1] Vignettes are short (and succinct) examples or situations of teaching actions. These can be presented to persons being tested in both text-based and video-based form for the purpose of content analysis, and they approximate real everyday classroom conditions. Persons being tested can take a position on them, as well as develop proposed solutions. Particular classroom situations are understood here as problem situations with respect to instructional design and teaching contents.

3.2. Research Process

Following the development of a theoretical competence structure model and the generation of vignettes including response-items (Figure 2), a first pilot study with students (N = 36) took place, as well as a validation of the vignettes and items using a three-stage expert survey design. To start with, all 30 teaching situations were qualitatively tested for content validity (*technical-didactic relevance, proximity to everyday classroom conditions*) in individual interviews (N = 8) with departmental heads from the technical division of public colleges of education and teacher training in *Baden-Württemberg* (Goreth, 2015; Jenßen, Dunekacke and Blömeke, 2015).

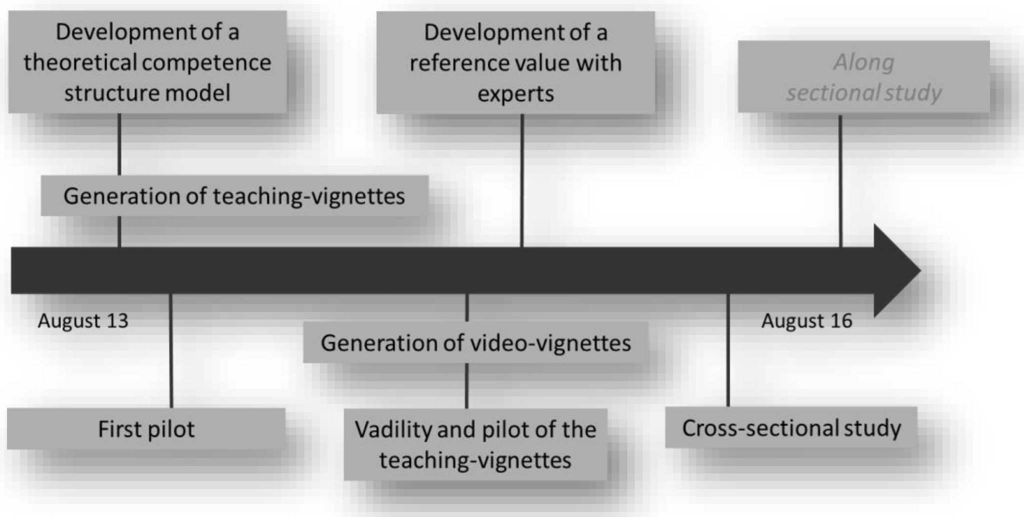

Fig. 2 - Test instrument development in the EKoL7 research project

For the purpose of developing a reference value within the teaching vignettes, a technical education expert norm (PCK-T) was determined. This norm is meant to depict what action is adequate in the classroom situation shown. To this end, we use a two-stage quantitative expert survey (Figure 3) guided by the Delphi method[2]. The survey panel consists of persons from different domains of expertise: specialised subject experts at schools of education and universities, departmental heads at public colleges for education and teacher training[3], as well as experienced teachers of technology (departmental advisors, college superintendents, teachers involved in training students doing internships).

For the purpose of identifying a generally valid norm of expert assessments of the generated teaching situations on which the procedure is based, the revised vignettes deriving from survey step I are placed online using the *soscisurvey.de* software, in order then to have them quantitatively evaluated by experts. Whereas the quality of the test-items are determined in the following survey step II ($N_2 = 79$), in order then to modify the test instrument and possibly to make a selection of vignettes or response-items, survey step III ($N_3 = 76$) serves to generate the expert norm, in order to be able undertake a comparative investigation of evaluations of teachers in training concerning the identification of adequate modes of action and of evaluations of the teaching situations represented (Tepner & Dollny, 2014).

These survey steps take place within a standardised framework, whereby, before the final survey step, the participants receive a brief notification and the modified survey instrument is submitted to them. Due to a partial modification of the test instrument, however, no statistical values are established here. In what follows, we present the quantitatively-designed expert survey steps and explain the issues deriving from them.

3.3. Problem and Research Questions

The goal is to develop a standardised reference value for the vignette-based closed-ended test format, which brings us to the problem addressed by this paper. The *heterogeneity of the sample* and hence the *differing origins* and *differing educational positions* of the surveyed experts are taken into account here:

- *Can an expert norm be developed as a reference value by using expert groups from different institutions?*

We expect only limited, non-significant differences in the evaluation of the teaching situations between the individual expert groups.

[2] As will be described in what follows: without feedback.
[3] Often called director of studies in *federal states* apart from Baden-Württemberg.

- *Do the different technical-education approaches (Schmayl, 2003) of the different federal states give rise to divergent evaluations?*
- *Are there differences in the evaluation of the classroom situations between expert groups from different federal states?*

We presume that there will be different evaluations of the teaching situations due to the different origin of the expert groups (different *federal states*), which have different theoretical backgrounds. These differences should not become significant, however, since, in our view, a different theoretical background has an effect more on divergent overall conceptions and less on the teaching segments.

- *Do age and gender have an impact on the evaluation of the teaching situations?*

It is not expected that gender- and age-specific differences will have any influence on the evaluation of the teaching situations anchored in the test.

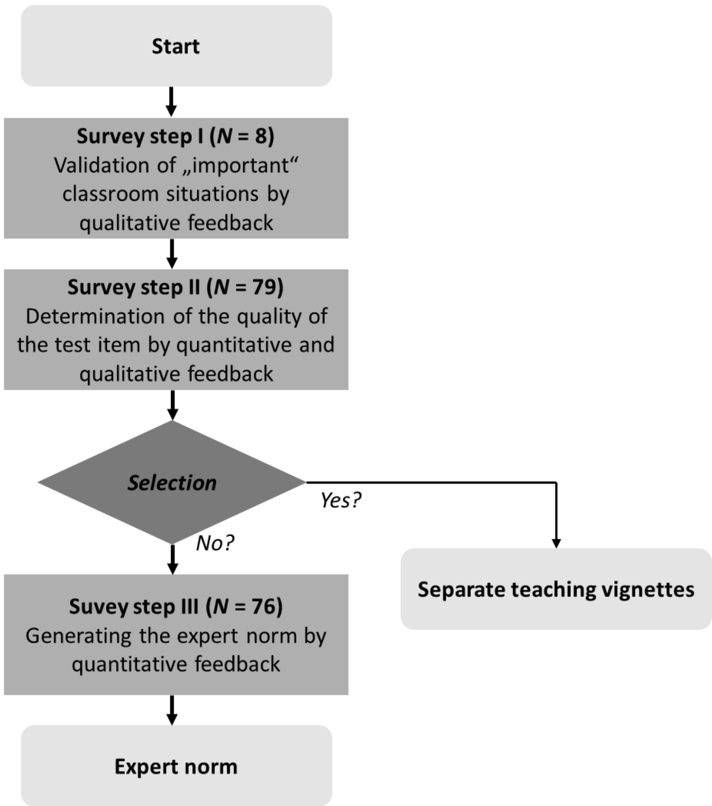

Fig. 3 - Survey steps in the development of the expert norm.

4. Results

4.1. Sample and Response

The electronically contacted sample consisted of N = 213 persons. The respondents are primarily either practitioners or professionals from universities, and within the school context, these can be regarded as the most important pillars of teacher training (Beutel, Gröschner & Lütgert, 2006). N = 79 persons (technical education experts) participated in the online survey. The survey was divided into three areas (*1. handling of tools and machines, 2. dealing with student conceptions and models in technology classes* and *3. dealing with methods in technology classes*), so that respondents selected the area with the highest self-ascribed degree of expertise. Moreover, additional areas can also be gone into after the approximately 1-hour duration of the survey.

Altogether, N = 76 persons took part in the third expert survey. The response rate was enhanced as described by Dillman, Smyth and Christian (2013). Hence, in order to increase the response rate, two reminders were sent during the approximately one-month duration of the survey (Dillman, Smyth & Christian, 2013).

Table 1 - Distribution of the sample for generating experts.

	Expert Survey II	Expert Survey III
Contacted experts	213	209
Retrieval of questionnaire	118 (= 55%)	110 (= 53%)
At least five items answered	79 (= 37%)	76 (= 36%)
Completed	66 (= 31%)	71 (= 34%)

Altogether, there was a 37% response rate in Expert Survey II (Table 1) and 36% in Expert Survey III. For an online survey, this represents a satisfactory response (Bolte 2003). In addition to the external criterion of expertise provided by holding a position in education, subjective confidence in judgment was surveyed using a six-point self-rating scale (1 = "not at all" to 6 = "very"). At M_{Sub} = 5.13 (SD = .86), this is very high across all vignettes and can be interpreted as further indication of a stable degree of expertise. From among the participating experts[4], 50% had passed the second State Examination [*Staatsexam*], 24% had a *Diplom* (roughly equivalent to a bachelor's/master's degree) and 22% had a doctorate/PhD[5].

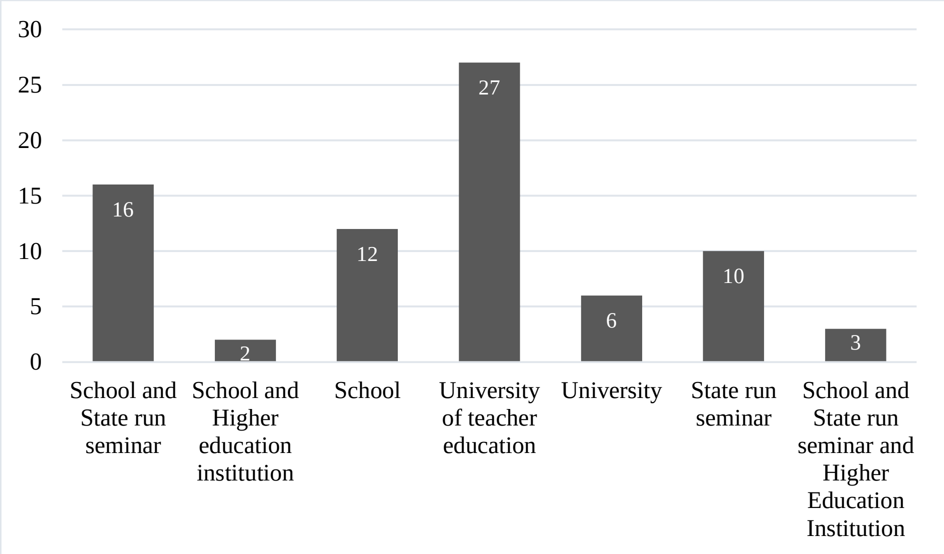

Fig. 4 - Sample distribution according to institution (N_3 = 74).

The sample is from different institutions[6] (higher education institutions: 38%; public teacher training colleges: 29%; schools: 33%¸Fig. 4). Overall, 63 males (= 85%) and 11 females (= 15%) participated. The respondents exhibit an average age of M = 46.9 (SD = 10.6; min. = 26; max. = 70; N_3 = 74). They live in various German *federal states*, as well as from Austria and Switzerland (see Table 2).

Table 2 - Place of work of surveyed experts by state or country of origin.

Place of work	Participating experts	Percentage
Baden-Württemberg	39	.54
Other Bundesländer (Lower Saxony, North Rhine-Westphalia, Schleswig-Holstein, Thuringia)	16	.22
Switzerland	5	.07
Austria	12	.17

4.2 Test Instrument Validation Steps

The data from Expert Survey II serves to establish the quality standard of the developed test item, in order to be able to use derived validation steps for the test instrument. For the selection of the closed-ended response items, the expert-item

[4] Multiple answers possible.
[5] As well as other answers that are not listed, like, for example: Magister: 15%; master's: 9%; bachelor's: 11% (N_3 = 74).
[6] In Germany the *pre-service Teacher Training* includes two steps: Step I at the *Higher Education Institutions* (theoretical) and Step II at the *Public Teacher Training Colleges* (practical).

matrices of each teaching situation are transposed and all experts with a correlation value of $r_{it} < .30$ are excluded from the vignette in question. This allows us to ensure an expert norm that is as consistent as possible. The selection of the items takes place according to the following criteria:

- unique modal value of an item (Figure 5: *Item 4 not bimodal*)
- standard deviation SD < 1.50 (Figure 5: *Item 1*)
- deviation from the modal value > 2 of less than 25% of the surveyed experts (Figure 5: *Item 3*).

	Item 1	*Item 2*	*Item 3*	*Item 4*	*Item 5*	*Item 6*	*Item 7*
Evaluate the following reactions/actions of the teacher in relation to the technical experiment.	**The teacher should present the best results in plenary.**	**The teacher should pull the phase of preparation into the break.**	**The teacher should include a lesson that sheds light on the subject.**	**The teacher should give up the hypothesis test as homework and have the results compared in the next lesson.**	**The teacher should draw a prepared solution as a pattern on the blackboard and have it written down.**	**The teacher should have the hypothesis review compared in the coming lesson.**	**The teacher should finsih the lesson on time and continue work in the following lesson.**
1 = *Strongly disagree*	IIII	IIIIIIIIIIIIII	IIIIII	IIII	IIIIIIIIIIIII	I	
2	I	IIIIIIII	IIIII	IIIII	IIIIIII		I
3	III	I	IIIII	II	III	II	I
4	IIII	I	IIII	III		IIII	IIIIII
5	IIIIIIIII		III	IIIIII		IIIIIIIIII	IIIIIII
6 = *Totally agree*	III		I	IIII	I	IIIIIII	IIIIIIIII
Modal	**5**	**1**	**1**		**1**	**5**	**6**
Standard deviation	**1.67**	**0.78**	**1,52**	**1.82**	**1.15**	**1.22**	**1.10**
N	**24**	**24**	**24**	**24**	**24**	**24**	**24**
Deviation	**21%**	**4%**	**33%**		**4%**	**4%**	**8%**

Fig. 5 - Item characteristics of the text vignette stability of paper bridges (n = 24 experts).

In addition, expert comments collected in an open-ended response field are also taken into account, in order to select responses or to modify response-items. Items with different responses (modal values) are preferred. Besides the item selection, a selection on the vignette level also takes place. To this end, those vignettes are removed from the test instrument that are not satisfactory according to the following criteria:

- Cronbach's Alpha below $\alpha = .85$ (accordance of experts of the transposed matrix)
- share of chosen experts over 30% within a vignette
- low content validity in terms of "technical-didactic relevance," "clarity of the teaching situation" and "clearly described situation"
- selection of qualitative comments.

The following figure (Figure 6) shows the theoretical structure model, including the developed teaching vignettes associated with it. Following the selection steps on both the item and the vignette level, 11 of the 30 teaching situations originally developed[7] are removed from the test.

Meanwhile, operationalised by way of *technical-didactic relevance*, *proximity to everyday classroom conditions*, and *the clarity of the teaching situation presented*, the identified content validity can be taken as given (Figure 7). High values, significantly above the middle value in the scale, can be recorded across all 30 vignettes[8]. Whereas the *technical-didactic relevance* of the depicted situations is estimated at $M_{TDR} = 4.96$ (SD = 1.12) by the surveyed experts, the values with respect to *proximity to everyday conditions* $M_{PEC} = 4.83$ (SD = 1.08) and the *clarity* of the vignettes $M_{Clar} = 4.53$ (SD = 1.41; $N_2 = 79$) are slightly lower.

[7] Including the separate vignettes *Hammer Use I* and *Hammer Use II*.

[8] Two vignettes have been merged.

Fig. 6 - Overview of vignette selection following survey step II (Goreth et al., 2016, p. 48).

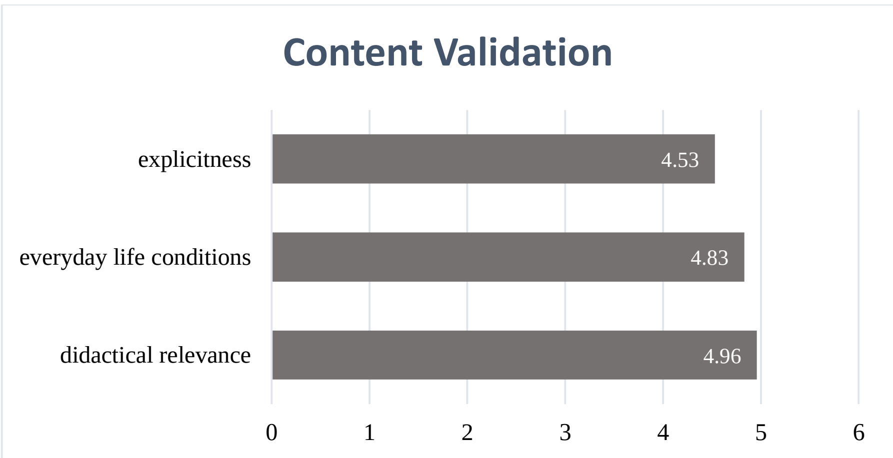

Fig. 7 - Content validation of the teaching situations presented (*clarity*; *proximity to everyday conditions*; *technical-didactic relevance* on a 6-point rating scale (1 = "not at all" to 6 = "very"; N_2 = 79).

If we consider the content validity of those teaching vignettes remaining in the final test instrument, on a 6-point scale (6 = very accurate), very high values are apparent precisely with respect to *technical-didactic relevance* (M_{TDR} = 5.16, SD = 1.02) and *proximity to everyday classroom conditions* (M_{PEC} = 5.07; SD = 1.01). The aspects evaluated by the experts can thus be regarded as positive for the testing of the content validity.

4.2. Expert Agreement

Given the different institutional domains of the persons involved in teacher training, in what follows, we will consider the question of whether a technical education expert norm can be produced by experts from higher education institutions, public teacher training colleges, and schools. Agreement within the teaching vignettes is identified by way of the internal consistency measure Cronbach's α. Whereas all the teaching situations developed obtain a level of agreement between $.69 < \alpha < .97$ within Expert Survey II, this value could be raised even more, to $.86 < \alpha < .97$, after the selective validation steps (Table 3). When averaged, there is a raised level of expert agreement of $\alpha = .93$ with a 15% share of excluded experts as compared to $\alpha = .88$ and a 22% share of removed experts before the selection steps. The teaching situations, as modified according to defined selection criteria (expert agreement: $r_{it} < .30$; $SD > 1.50$; share of excluded experts < 25%), register an increased in expert agreement from $\alpha_{Surv2} = .93$ ($N_2 = 79$) to $\alpha_{Surv3} = .95$ ($N_3 = 76$). The test procedure covers relevant components of professional skills on the basis of 15 Vignettes (11 text-based; 4 in video form).

Table 3 - Expert agreement in survey steps II and III ($N_2 = 79$; $N_3 = 76$; Goreth et al., 2016, p. 52).

	Vignette	Exp.-Surv. II (α)	Excl. Exp. (in %)	Exp.-Surv. III (α)	Excl. Exp. (in %)
1	Woodworking	.97	.20	.96	.14
2	Correct Filing	.95	.08	.96	.05
3	The Robot	.90	.17	.97	.18
4	The Consumed Electricity	.91	.13	.93	.29
5	The Cable Cord	.97	.04	.99	.03
6	A File is Needed	.91	.20	.95	.14
7	Functional Analysis	.89	.28	.96	.19
8	Hammer Use	.94	.14	.85	.43
9	Insulation Attempt	.86	.29	.95	.23
10	Fast Preparation	.90	.15	.95	.24
11	Transformer Voltage	.94	.11	.91	.24
12	Stability Paper Bridges	.94	.20	.99	.16
13	Drill Press	.96	.05	.96	.04
14	Workpiece Secured	.90	.26	.96	.05
15	Tool Course	.93	.04	.96	.23
	Total	***.93***	***.15***	***.95***	***.18***
16	Electric Shock	.96	.04	*The test comprises 15 vignettes (teaching vignettes 16-29 were excluded and vignette 8* was merged)*	
17	Key Questions	.82	.41		
18	Paper Bridges	.81	.33		
19	Backsaw	.80	.36		
20	Handsaw	.68	.35		
21	The Bell Circuitry	.93	.14		
22	Posing the Problem	.94	.17		
23	Moped Torque Increase	.69	.50		
24	Group Work – Construction	.72	.64		
25	The Otto Engine Model	.88	.28		
26	The Small Transformer	.85	.29		
27	The Power Grid	.95	.17		
28	Starting Out	.87	.23		
29	Threading	.81	.39		
	Total	***.88***	***0.22***		

In addition, however, there could be hidden differences, since groups of people (for instance, from higher education institutions) may evaluate the vignettes presented differently than groups from schools. Hence, in the following, we will examine the extent to which experts from *higher education institutions*, *public teacher training colleges*, *schools*, and *public teacher training colleges + schools*[9] differ from one other in their judgments. All 88 response-items remaining in the test are examined for significant differences in means using a t-test. Only one item can be found that exhibits a

[9] For the exact group composition, Fig. 4.

significant difference in evaluation between the *schools* and *public teacher training colleges* groups, as well as the *schools* and *public teacher training colleges and schools* groups. (This corresponds to an agreement between the judgments given of 98.8%; Table 4). Similarly, only three items exhibit a significant difference in evaluation between the *schools* and *higher education institutions* groups. (This corresponds to a rate of agreement of 96.5%) Overall, only a few items are differently assessed between the institutions of the respondents. This means that the underlying institutions only have a minimal influence on the response to the teaching situations presented and that it is possible to develop an expert norm on the basis of people from the domains in question.

Table 4 - Agreement of expert judgments by institutional group characteristic (N_3 = 76[10]; Goreth et al., 2016, p. 54).

Group Comparison	Number of Matches in the Items (total: 88 items)
Schools and Public Colleges + Schools	87 (= 98.8%)
Schools and Higher Education Institutions	85 (= 96.5%)
Schools and Public Colleges	87 (= 98.8%)

Furthermore, it would be intriguing to examine whether an expert norm can be generated on the basis of the different normative technical-didactic approaches in the Federal Republic of Germany (Schmayl, 2003) or whether these differences lead to differences in the evaluation of the teaching vignettes presented. Whether the age or gender of the respondents leads to divergent expert judgments is also of interest. To this end, the response-items (mean values) in the individual question-items will be examined for significant differences in relation to place of work (BW – not BW), as well as, in addition, with regard to age and gender. The different groups (G_1: BW – not BW; G_2: young in age – advanced age; G3: female – male) are compared with each other. In making the comparison, all the individual response-items are analysed with respect to their mean-value differences, taking into account the respective expert background. If the groups to be examined (for example, gender: male – female) do not agree in their evaluation of an item on a six-point rating scale (Wirtz, 2020), it can be presumed that the group characteristic has an influence on how the response-item is answered. Hence, for the purpose of a complete assessment, all the items compiled are examined for significant mean-value differences using a t-test.

Table 5 - Mean-value comparisons by group characteristics place of work[11], age and gender (N_3 = 76; Goreth et al., 2016, p. 54).

Group Comparison	Teaching Structure in % (42 items)	Handling of Tools and Machines in % (46 items)	Total in % (88 items)
Place of Work (BW – not BW)	.90	.93	.92
Age	.98	.91	.94
Gender	1.00	.98	.99

Whereas place of work (BW – not BW), with 8% significant differences in means, has a minimal influence on the answering of the response-items, in the case of the group characteristics age[12], with overall 94% of answered items matching, and gender, with only 1% of the teaching items judged differently (Table 5), these values are likewise low. The group characteristics examined (place of work, age and gender) also exhibit only a marginal influence on the evaluation of the teaching vignettes. Persons from the different domains of teacher training and development can be drawn upon for the development of a reference value. They generate a high level of agreement in the teaching situations and can be used in the further conduct of the study.

5. Discussion and Conclusion

5.1. Discussion

The facets of professional performance of teachers that are deemed important in the current literature were taken up in the present paper. Whereas the models of Voss et al. (2015) and Baumert and Kunter (2013) focus on the aspects of PCK, (G)PK and CK, the *professional perception of teaching* represents an expansion of the model of the teaching process (Meschede et al., 2015).

[10] The number of matches is calculated from the quotient of the significant differences in means (t-test) and the total number of items.
[11] The number of matches is calculated from the quotient of the significant mean differences (t-test) and the number of items.
[12] Comparatively considered, two groups were divided according to the median.

Vignette-based teaching research from the neighbouring areas of mathematics and natural science teaching can largely be drawn upon for the present project. As in previous projects, a reference value was determined by way of experts whose performance in their area is outstanding and who have a wealth of experience. This reference value can be used for a comparison of student responses. Tepner and Dollny (2014) or Seidel, Blomberg and Stürmer (2010), for instance, provide guidance in this regard.

In our own project, to begin with, all the developed teaching vignettes were tested for content validity with departmental heads in a multi-stage expert survey process and the teaching situations were then modified. Next, experts from the institutional domains of schools and of universities, as well as teacher training colleges and schools, were drawn upon in a quantitative survey structure. It turned out that a technical education expert norm (PCK-T) could be established on the basis of different groups. A satisfactory response of 37% (N = 79) and 39% (N = 76; Bolte, 2003) allowed data to be generated that serves, on the one hand, to modify and select (according to definite criteria) the test instrument. On the other hand, it could be shown that both the content validity of the final test instrument, comprising 15 vignettes, represents an adequate content validity (*technical-didactic relevance*: M = 5.16, SD = 1.02 and *proximity to everyday teaching conditions*: M = 5.07; SD = 1.01; N = 79 on a 6-point rating scale) and, moreover, that a very high level of expert agreement is achieved. Following revision, as well as following defined selection steps, the latter could be raised from $\alpha = .93$ to $\alpha = .95$ in the final survey step. An examination according to certain characteristics (*institution*, *place of work*, *age* and *gender*) could only provide marginal influences on the evaluation of the closed-ended response-items.

5.2. Conclusion

The following can be concluded: An examination according to certain characteristics (institution, place of work, age and gender) could only have a marginal influence on the evaluation of the closed response items. The determination of a reference value using a multi-stage expert survey structure and generated from groups of persons in different institutions thus appears possible, and it can be used in the further course of the project to consider student responses in a comparative perspective.

This norm (PCK-T) is the result of both qualitative and quantitative parts, even if, in the present investigative structure, it is not possible to draw any conclusions about the background to the experts' evaluation. Furthermore, data from a cross-sectional study of students (N = 350) is used, in order to undertake further tests of validity and to investigate possible effects due to existing study subject combinations (Goreth 2017).

Acknowledgement

Special thanks to all participants and institutions that were involved in this study.

References

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell (Ed.), *Handbook of research on science education* (1105-1149). Mahwa, NJ: Lawerence Erlbaum Associates.

Atteslander, P. (2010). *Methoden der empirischen Sozialforschung* (13th ed.). Berlin: Erich Schmidt.

Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COAKTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COAKTIV* (29-53). Münster: Waxmann.

Baumert J. & Kunter M. (2013). Professionelle Kompetenz von Lehrkräften. In I. Gogolin, H. Kuper, H. H. Krüger & J. Baumert (Eds.), *Stichwort: Zeitschrift für Erziehungswissenschaft* (277–337). Wiesbaden: Springer.

Beutel, S.-I., Gröschner, A. & Lütgert, W. (Eds.). (2006). *Lehrerbildung im Wandel. Eine Expertenbefragung zu den Perspektiven einer inhaltlichen und strukturellen Reform*. Jena: edition paideia.

Blomberg, G., Sherin, M. G., Renkl, A., Glogger, I., & Seidel, T. (2014). Understanding video as a tool for teacher education: investigating instructional strategies to promote reflection. *Instructional Science*, 42, 3. 443–463.

Bolte, C. (2003). Konturen wünschenswerter chemiebezogener Bildung im Meinungsbild einer ausgewählten Öffentlichkeit - Methode und Konzeption der curricularen Delphi-Studie Chemie sowie Ergebnisse aus dem ersten Untersuchungsabschnitt. *Zeitschrift für Didaktik der Naturwissenschaften*, 9. 7-26.

Brovelli, D., Bölsterli, K., Rehm, M. & Wilhelm, M. (2013). Professionelle Kompetenz und Berufsidentität in integrierten und disziplinären Lehramtsstudiengängen der Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 17. 57-87.

Bruckmaier, G., & Krauss, S. (2015). Modellieren in der COACTIV-Videostudie. In G. Kaiser & H. Henn (Eds.), *EBL-Schweitzer. Werner Blum und seine Beiträge zum Modellieren im Mathematikunterricht. Festschrift zum 70. Geburtstag von Werner Blum* (77-88). Wiesbaden: Springer.

Dillman, D. A., Smyth, J. D. & Christian, L. M. (2014*). Internet, phone, mail and mixed-mode surveys. The tailored design method* (4th ed.). Hoboken, N.J.: Wiley.

Frey, A. & Jung, C. (2011). Kompetenzmodelle und Standards in Lehrerbildung und Lehrberuf. In E. Terhart, H. Bennenwitz & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrberuf* (540-572). Münster: Waxmann.

Goreth, S. (2017). Erfassung und Modellierung professioneller Unterrichtswahrnehmung angehender Lehrkräfte im naturwissenschafts- und technikbezogenen Unterricht. Dissertation. In B. Geißel & T. Gschwendtner (Eds.), *Beiträge zur Technikdidaktik*, 4. Berlin: logos.

Goreth, S., Schray, H., Rehm, M. & Geißel, B. (2016). Die Entwicklung einer technikdidaktischen Expertennorm als Referenzkriterium zur Auswertung von Unterrichtsvignetten. In B. Geißel & T. Gschwendtner (Eds.), *Beiträge zur Technikdidaktik*, 1. Aktuelle Forschungsarbeiten und unterrichtspraktische Beispiele (33-61). Berlin: logos.

Goreth, S. (2015). Testinstrumententwicklung zur Erfassung fachdidaktischer Lehrkompetenz im technikbezogenen Unterricht der Sekundarstufe. In W. Bienhaus & C. Wiesmüller (Eds.), *Technische Bildung und MINT. Chance oder Risiko* (173-187). Weingarten: Deutsche Gesellschaft für technische Bildung.

Goreth, S., Geißel, B. & Rehm, M. (2015). Erfassung fachdidaktischer Lehrkompetenz im technikbezogenen Unterricht der Sekundarstufe 1. - Instrumentenkonstruktion und erste Befunde. *Journal of Technical Education,* 3, 1. 13-38.

Gruber, H. & Mandl, H. (1996): Expertise und Erfahrung. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung. Theoretische und methodische Grundlagen*. Opladen: Westdeutscher Verlag.

Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3rd ed.). Wiesbaden: Springer.

Häder, M. (2009). *Delphi-Befragungen. Ein Arbeitsbuch* (2nd ed.). Wiesbaden: Verlag für Sozialwissenschaften.

Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren* (3rd ed.). Stuttgart: Kohldammer.

Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K. & Klieme, E. (2009). Teaching patterns and learning quality in swiss and german mathematics lessons. *Learning and Instruction*, 29, 1. 66-78.

Jenßen, L., Dunekacke, S. & Blömeke S. (2015). Qualitätssicherung in der Kompetenzforschung. *Z Päd*, 61. 11-31.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments: Theoretical & Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, 13, 2. 369–387.

KMK (2015). *Übersicht Lehrpläne: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK)*. Available online at http://www.kmk.org/dokumentation/lehrplaene/uebersicht-lehrplaene.html, last accessed on 10/15/2015.

König, J. (2015). Kontextualisierte Erfassung von Lehrerkompetenzen. *Zeitschrift für Pädagogik*, 61, 3. 305–309.

Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften –Methodische Herausforderungen. *Unterrichtswissenschaft*, 38, 1. 68-86.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2013). *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers. Results from the COACTIV Project*. New York et al.: Springer.

Lehmann, G. (2012). Wertorientiertes Management mittelständischer Unternehmungen. Potentiale der Ermittlung von Eigenkapitalkosten sowie Herleitung von Konzeptanforderungen im Rahmen eines Delphi-Verfahrens. In C. Steinle (Ed.), *Schriften zum Management*. München: Rainer Hampp.

Meschede, N., Steffensky, M., Wolters, M., & Möller, K. (2015). Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht: Theoretische Beschreibung und empirische Erfassung Professional Vision on Science Lessons in Primary School - Conceptualization and Measurement. *Unterrichtswissenschaft*, 43, 4. 317-335.

Meschede, N. (2014). Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht: Theoretische Beschreibung und empirische Erfassung. In H. Niedderer, H. Fischer & E. Sumfleth (Eds.), *Studien zum Physik- und Chemielernen. Band 163*. Berlin: Logos.

Oser, F. (2015). Wer setzt das Maß? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz. *Zeitschrift für Pädagogik*, 61, 3. 361-376.

Posner, M. I. (1988). Introduction: What is it to be an expert? In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (xxix-xxxvi). Hillsdale, NJ: Erlbaum.

Rehm, M., & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (213–225). Berlin, Heidelberg: Springer.

Rehm, M., Bünder, W., Haas, T., Buck, P., Labudde, P., Brovelli, D., Ostergaard, e., Rittersbacher, C., Wilhelm, M., Genseberger, R. & Svoboda, G. (2008). Legitimierungen und Fundamente eines integrierten Unterrichtsfachs "Science". *Zeitschrift für Didaktik der Naturwissenschaften*, 14. 99-124.

Riegel, U. (2013). Videobasierte Kompetenzforschung in den Fachdidaktiken. Einleitung. In U. Riegel & K. Macha (Eds.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (9-24). Münster et al.: Waxmann.

Riese, J., Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Schecker, H., Tomczyszyn; E. & Zander, S. (2015) Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. *Zeitschrift für Pädagogik*, 61. 55-79.

Rohaan, E., Taconis, R. & Jochems, W. (2012). Analysing teacher knowledge for technology education in primary schools. *Int J Technol Des Educ*, 22. 271–280.

Rohaan, E. (2009). *Testing Teacher Knowledge for Technology Teaching in Primary Schools*. Eindhoven University of Technology Library: Printservice TU/e.

Schmayl, W. (2003). Ansätze allgemeinbildenden Technikunterrichts. In B. Bonz & B. Ott (Eds.), *Allgemeine Technikdidaktik – Theorieansätze und Praxisbezüge. Berufsbildung konkret* (131-147). Baltmannsweiler: Schneider.

Seidel, T., Blomberg, G. & Stürmer, S. (2010). „Observer" – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Eckhard, D. Leutner & M. Kenk (eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. *Zeitschrift für Pädagogik*, 56 (296-306). Weinheim/Basel: Beltz.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57. 1-21.

Steffensky, M., Gold, B., Holdynski, M., & Möller, K. (2015). Professional Vision of Classroom Management and Learning Support in Science Classrooms-Does Professional Vision Differ Across General and Content-Specific Classroom Interactions? *International Journal of Science and Mathematics Education*, 13, 2. 351-368.

Stender, A. (2014). *Unterrichtsplanung: Vom Wissen zum Handeln: Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung. Studien zum Physik- und Chemielernen*. Berlin: Logos.

Sternberg, R. J. & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, 24, 6. 9-17.

Stürmer K. & Seidel T. (2015). Assessing Professional Vision in Teacher Candidates. *Zeitschrift für Psychologie*, 223, 1. 54-61.

Tenberg, R. (2014). Competence-oriented study – didactic reform of higher education or Bologna-Rhetoric? *Journal of Technical Education*, 2, 1. 1-15.

Tenberg, R. (2016). How can we get technology taught in schools? *Journal of Technical Education*, 4, 1. 1-10.

Tepner, O. & Dollny, S. (2014). Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens. In D. Krüger (Ed.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (311–323). Berlin: Springer.

Tepner, O., Borowski, A., Dollny, S., Fischer, H., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B., Sandmann, A., Sumfleth, E., Thillmann, H. & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 7-28.

Traebert, W.-E. (2001). Die Bedeutung der Praxis für den Technikunterricht. In B. Sachs & C. Sachs (Eds.), *Praxis und Theorie in der Technischen Bildung. 4. Tagung der DGTB in Wilhelmshaven vom 14.-16.9.2000* (12-21). Villingen-Schwenningen: Neckar-Verlag.

VDI (2007). *Bildungsstandards Technik für den mittleren Bildungsabschluss*. Düsseldorf: VDI.

VDI (2012). Technische Allgemeinbildung stärkt den Standort Deutschland. Positionspapier des Vereins Deutscher Ingenieure (VDI). *tu – Zeitschrift für Technik im Unterricht*, 41, 146. 13.

Voss, T., Kunina-Habenicht, O., Hoehne, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18, 2. 187–223.

Voss, T., Kunter, M., & Baumert, J. (2011). Assesing Teacher Candidates. *General Pedagogical/Psychological Knowledge: Test Construction and Validation*, 103. 952–969.

Wirtz, M. A. (Eds.). (2020). Dorsch - Lexikon der Psychologie (19. edition). Göttingen: Hogrefe.

Witner, S. & Tepner, O. (2011). Entwicklung geschlossener Testaufgaben zur Erhebung des fachdidaktischen Wissens von Chemielehrkräften. *Chimica didacticae: Zeitschrift für die Didaktik aller Naturwissenschaften*, 113-137.

Zinn, B. (2014). Technological literacy – Relevance spectrum, educational standards and research perspectives. *Journal of Technical Education*, 2, 2. 1-23.