International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge

Sophie Kirschner, Andreas Borowski, Hans E. Fischer, Julie Gess-Newsome & Claudia von Aufschnaiter

To cite this article: Sophie Kirschner, Andreas Borowski, Hans E. Fischer, Julie Gess-Newsome & Claudia von Aufschnaiter (2016) Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge, International Journal of Science Education, 38:8, 1343-1372, DOI: 10.1080/09500693.2016.1190479

To link to this article: https://doi.org/10.1080/09500693.2016.1190479

Published online: 14 Jun 2016.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
VOL. 38, NO. 8, 1343–1372
http://dx.doi.org/10.1080/09500693.2016.1190479

Routledge
Taylor & Francis Group

Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge

Sophie Kirschner[a], Andreas Borowski[b], Hans E. Fischer[c], Julie Gess-Newsome[d] and Claudia von Aufschnaiter[a]

[a]Physics Education, Justus Liebig University Giessen, Giessen, Germany; [b]Physics Education, University of Potsdam, Potsdam-Golm, Germany; [c]Physics Education, University Duisburg-Essen, Essen, Germany; [d]College of Education, Oregon State University-Cascades, Bend, OR, USA

ABSTRACT

Teachers' professional knowledge is assumed to be a key variable for effective teaching. As teacher education has the goal to enhance professional knowledge of current and future teachers, this knowledge should be described and assessed. Nevertheless, only a limited number of studies quantitatively measures physics teachers' professional knowledge. The study reported in this paper was part of a bigger project with the broader goal of understanding teacher professional knowledge. We designed a test instrument to assess the professional knowledge of physics teachers ($N = 186$) in the dimensions of *content knowledge* (CK), *pedagogical content knowledge* (PCK), and *pedagogical knowledge* (PK). A model describing the relationships between these three dimensions of professional knowledge was created to inform the design of the tests used to measure CK, PCK, and PK. In this paper, we describe the model with particular emphasis on the PCK part, and the subsequent PCK test development and its implementation in detail. We report different approaches to evaluate the PCK test, including the description of content validity, the examination of the internal structure of professional knowledge, and the analysis of construct validity by testing teachers across different school subjects, teachers from different school types, pre-service teachers, and physicists. Our findings demonstrate that our PCK test results could distinguish physics teachers from the other groups tested. The PCK test results could not be explained by teachers' CK or PK, cognitive abilities, computational skills, or science knowledge.

ARTICLE HISTORY
Received 25 March 2015
Accepted 12 May 2016

KEYWORDS
Pedagogical content knowledge; physics education; teacher knowledge; quantitative research

Teachers' professional knowledge has been recognized as an essential factor for effective teaching (Abell, 2007). Typically, *content knowledge* (CK), *pedagogical content knowledge* (PCK), and *pedagogical knowledge* (PK) are described as dimensions of professional knowledge. However, only a limited number of studies actually quantitatively measure

CONTACT Sophie Kirschner sophie.kirschner@didaktik.physik.uni-giessen.de Justus Liebig University Giessen, Karl-Gloeckner-Strasse 21 C, 35394 Giessen, Germany

teachers' knowledge (Abell, 2008; Smith & Banilower, 2015). Whereas researchers in mathematics education are pioneers in using large scale assessments to analyze pre-service and in-service teachers' knowledge (Baumert et al., 2010; Blömeke, Felbrich, Müller, Kaiser, & Lehmann, 2008; Hill, Rowan, & Ball, 2005), only rarely is professional knowledge of science teachers assessed (some exceptions are, for instance, Daehler, Heller, & Wong, 2015; Jüttner & Neuhaus, 2012; Olszewski, 2010). The study reported in this paper is part of the ProwiN project ('Professional Knowledge in Science') with the broader goal of understanding teacher professional knowledge. ProwiN research is characterized by systematic and model-based test development (Tepner et al., 2012) and has resulted in tests for biology, chemistry, and physics teachers (Borowski et al., 2010). In this paper, we describe the general model of the relationships between the dimensions of professional knowledge as it is applied to physics. We explain how the instrument was tailored to the model by providing, classifying, and discussing example items with a focus on content validity. Then, we report how the internal structure of the instrument was confirmed with test data and construct validity was analyzed through different approaches.

Theory

An important contribution to the description of teachers' professional knowledge has been Shulman's introduction of PCK as 'subject matter knowledge for teaching' (Shulman, 1986, p. 9) and his suggestion that professional knowledge has seven aspects (1987, p. 8): (1) CK, (2) general PK, (3) curriculum knowledge, (4) PCK, (5) knowledge of learners and their characteristics, (6) knowledge of educational contexts, and (7) knowledge of educational ends, purposes, and values. In the years after his theory on teachers' professional knowledge, his distinctions have been modified several times (e.g. Bromme, 1994; Grossman, 1990; Magnusson, Krajcik, & Borko, 1999). Recent research seems to share the assumption that three different dimensions are important to describe teachers' professional knowledge, which includes most of Shulman's ideas (e.g. Abell, 2007; Baumert et al., 2010; Blömeke, Felbrich, et al., 2008; Gess-Newsome, 1999; Grossman, 1990):

- CK, which comprises subject matter knowledge of phenomena, rules, theories, models, and their relationships necessary for teaching;
- PK, which comprises aspects about teaching and learning that are subject-independent;
- PCK, which comprises knowledge of teaching and learning of a particular subject and can be considered a synthesis of content and school related pedagogy (Shulman, 1987).

Another element of teachers' professional knowledge to consider is contextual knowledge: teachers have to 'adapt their more general knowledge to specific school settings and individual students' (Grossman, 1990, p. 9). In addition to teachers' professional knowledge, their beliefs are assumed to be part of teachers' competency (Blömeke, Felbrich, et al., 2008) and act as an important foundation of teachers' decisions (Fletcher & Luft, 2011, p. 1125). However, this study did not address contextual knowledge or teachers beliefs. This paper focuses on PCK which is seen as the aspect of teachers' professional knowledge that distinguishes subject-specific teachers from scientists on the one hand

and from pedagogues on the other hand. This distinction is necessary to describe PCK as specific for teachers of certain subjects, in this case, physics teachers.

Even though researchers agree on the major dimensions of teachers' professional knowledge and about the relevance of PCK (see Abell, 2007, 2008; Baumert et al., 2010), PCK has been conceptualized in different ways, partly because research aims differ. Some research aims to measure PCK, whereas other research aims to capture PCK (Park & Suh, 2015). Studies measuring PCK in science education (e.g. Daehler et al., 2015; McNeill, González-Howard, Katsh-Singer, & Loper, 2015; Schmelzing et al., 2013) and in math education (e.g. Baumert et al., 2010; Hill et al., 2005) address PCK on action (the knowledge of, reasoning behind, and planning for teaching). These studies do not address the act of teaching (PCK in action, Gess-Newsome, 2015, p. 36). They aim to develop efficient instruments for the long-term objective of gaining generalizable information about the connection between professional knowledge and student outcomes with large sample sizes. Typically, quantitative instruments, often paper-and-pencil tests, are used in these studies. Studies which aim to capture PCK (e.g. Alonzo & Kim, 2015; Friedrichsen, 2015; Park & Chen, 2012) address both: PCK in action and PCK on action. They often focus on the characteristics of PCK and its relationship to teaching and to analyzing instructional practice (Park & Suh, 2015). Because of the complexity, qualitative instruments – such as the analysis of observations (videos and protocols), interviews, or written tasks – are often used. Because of small sample sizes, researchers of these studies struggled to provide generalizable results.

The different approaches utilize various models. In order to develop instruments and procedures to measure professional knowledge quantitatively, models include a limited number of characteristics. In contrast, approaches capturing professional knowledge describe more comprehensive models. Furthermore, models can differ even though the research aims seems to be rather similar. A key difference between these models is the assumed interrelationship between CK, PCK, and PK. Many studies share the assumption of distinct dimensions for CK, PCK, and PK (e.g. Krauss et al., 2008; Magnusson et al., 1999; Park & Chen, 2012), whereas others consider CK and PK as parts of PCK (e.g. Hashweh, 2005; Loughran, Berry, & Mulhall, 2006; Rollnick, Bennett, Rhemtula, Dharsey, & Ndlovu, 2008). Table 1 lists different studies quantitatively measuring PCK and describing PCK on the basis of a qualitative study (capturing PCK) or a literature analysis. Listed in the columns 'knowledge about content' and 'knowledge about pedagogy' are studies where CK and PK are seen as facets of PCK (indicated with '+') or as independent dimensions (indicated with '–'). Furthermore, Table 1 gives an overview about facets typically used to characterize PCK and indicates which facet is considered in which study. The common ground for the facets lies in the Magnusson model (Magnusson et al., 1999). *Knowledge about student understanding* and *knowledge about instructional strategies and representations* are considered to be key facets (Alonzo & Kim, 2015; van Driel, Verloop, & de Vos, 1998; Park & Oliver, 2008; Park & Suh, 2015; Shulman, 1986). Enhancing student understanding is one main aim of teaching science. Knowledge about students is essential and serves as the foundation for instructional strategies and methods to achieve student understanding, as Grossman (1990) put it, 'To generate appropriate explanations and representations, teachers must have some knowledge about what students already know about a topic and what they are likely to find puzzling' (p. 8). Even though key facets can be identified in various models, it should be noted that different authors interpret similar

Table 1. Facets of PCK derived from different studies based on Park and Oliver (2008, p. 265).

		Facets: Knowledge about …							
Study	Kind of study	Content	Pedagogy	Context	Student understanding	Instructional strategies and representations	Curriculum	Purposes for teaching a subject	Assessment
Blömeke, Kaiser, and Lehmann (2008)	M	–	–		+	+	+		
Gramzow (2015)	M	–	–		+	+	(+)	(+)	(+)
Krauss et al. (2008)	M	–	–	–	+	+			–
Olszewski (2010)	M	–	–		+	+	+		
Schmelzing et al. (2013)	M	–	–		+	+			
Smith and Banilower (2015)	M	–	–		+	+	–	–	+
Tepner et al. (2012)	M	–	–		+	+			
Gess-Newsome, Carlson, Gardner, and Taylor (2010)	M	+	+	+	+	+			
Alonzo and Kim (2015)	D	–			+	+	–		–
Cochran, DeRuiter, and King (1993)	D	+	+	+	+				
Geddis (1993)	D				+	+	+		
Grossman (1990)	D	–	–	–	+	+	+	+	
Hashweh (2005)	D	+	+	+	+	+	+	+	+
Loughran et al. (2006)	D	+	+	+	+	+		+	
Magnusson et al. (1999)	D	–	–	–	+	+	+	+	+
Marks (1990)	D	+			+	+			
Park and Chen (2012)	D	–	–		+	+	+		+
Rollnick et al. (2008)	D	+	+	+	+	+	+		+
Smith and Neale (1989)	D	–			+	+		+	
Shulman (1987)	D	–	–	–	+	+	–	–	
Tamir (1988)	D	–	–		+	+	+		+
van Driel et al. (1998)	D	–			+	+			

Notes: M: Measuring PCK; D: Describing PCK on the basis of a qualitative study (capturing PCK) or on literature analysis; +: facet explicitly included; –: facet explicitly excluded; (+): facet included in model but excluded in the corresponding instrument.

facets differently. For example, knowledge about student understanding is sometimes interpreted as knowledge about students' alternative conceptions and sometimes as knowledge about learners and learning. The latter may include the former; however, these two interpretations of a specific facet cannot be regarded as identical, and therefore, comparability between different models of PCK is limited.

So far, we have argued that assumptions about the relationship between PCK and CK/PK differ, different facets are included in PCK by different researchers, and even if similar facets are identified, their meaning may vary. Studies measuring PCK often, but not always, treat PCK, CK, and PCK as distinct dimensions of professional knowledge (see Table 1: 'knowledge about content' and 'knowledge about pedagogy') and operationalize the facets 'knowledge about student understanding' and 'knowledge about instructional strategies and representations'.

The aim of our research was to develop a test instrument for assessing physics teachers' professional knowledge, to explore the validity of the PCK test, and to analyze how CK, PCK, and PK relate to each other. To achieve this aim, four steps were carried out:

(1) We modeled the relationships between CK, PCK, and PK theoretically.
(2) We established a model to operationalize physics teachers' professional knowledge and used it to develop a test instrument, ensuring content validity.
(3) In order to evaluate the structural validity of the test instrument, we investigated to find out if CK, PCK, and PK are distinct and reliable dimensions of teachers' professional knowledge and analyzed the correlations between these dimensions.
(4) In order to evaluate the construct validity of the test instrument, we tested groups that were assumed to differ in physics PCK.

In our description of teacher professional knowledge, we first discuss CK, PCK, and PK, with a special emphasis on PCK because we assumed that PCK is a component that is attributed to physics teachers only. Construct validity is discussed only for the PCK test developed in this study.

Model of professional knowledge

With limited consensus in the research community, a new model for teacher professional knowledge was developed in the ProwiN project, taking common characteristics into account. The ProwiN model describes the professional knowledge of biology, chemistry, and physics in-service teachers (Tepner et al., 2012) and is adapted and discussed in more detail for physics teachers in this paper. The model was required to structure test instrument development in our study. If a model describes a large number of different characteristics for each dimension, many items are required to cover the model. It follows that if larger numbers of items are used, either the time required for test completion has to be increased or the number of participants has to be high in order to distribute tasks using multimatrix sampling (Gonzalez & Rutkowski, 2010). Assuming that it would be difficult to recruit large samples of in-service teachers, we decided to focus on a limited number of characteristics that seem to be most relevant to the teaching profession. By limiting our model to a few variables, we were able to restrict our instrument in its coverage and take the first steps toward the development of a sound instrument. We

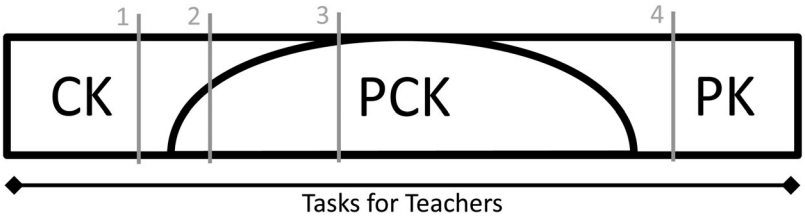


Figure 1. Model about the relationship between the dimensions of professional knowledge. The curved black line indicates that CK, PCK, and PK are distinct dimensions. The gray vertical lines indicate tasks that contain a different portion of the different dimensions of professional knowledge (line 1: CK only required for solution of the task, line 2: CK and PCK required at the same time, lines 3 and 4: PCK, respectively PK only required).

carefully describe here the theoretical model of teacher professional knowledge, focusing firstly on the relationships between CK, PCK, and PK, and secondly on the aspects used to operationalize professional knowledge. In our model, CK, PCK, and PK are regarded as distinct dimensions (see Figure 1, black curved line). As PCK is related to both CK and PK and this relationships should be closer than the one between CK and PK, we created a representation that has a linear layout (see Figure 1) ranging from CK (left) to PCK (middle) and PK (right). Problems and tasks are likely to require teachers to activate not only their PCK but also their CK or PK with varying levels of intensity, which is demonstrated by gray vertical lines in Figure 1. For example, 'Write down the lever principle' is a task which is associated clearly to CK (see Figure 1, vertical line 1). The question 'How can you change the experiment in a way that students can identify the relationship between the two variables?' requires both CK and PCK from teachers (see Figure 1, vertical line 2). An example for PCK only is 'Why is it important to consider students' preconceptions in lessons?' (see Figure 1, vertical line 3) and for PK only is 'What action is more adequate to stimulate self-regulated learning: Talk to the student or to his/her parents?' (see Figure 1, vertical line 4). The gradient of the lines between PCK and CK/PK in Figure 1 is representational and the portions of PCK and CK/PK areas cut through by the vertical lines for specific tasks are illustrative only.

Model to operationalize physics teachers' professional knowledge

The graphical representation in Figure 1 is helpful for establishing the relationships between the dimensions of professional knowledge. However, it reveals little about the dimensions themselves. In order to develop appropriate tasks for the assessment, characteristics of the dimensions need to be defined. For each of the dimensions of professional knowledge – PCK, CK, and PK – our model elaborates on the following characteristics: the *knowledge areas* covered, the physics *topics* on which we focused, and the *facets* included (see Figure 2). The following section describes each of these characteristics in detail.

Knowledge areas

For teaching, a teacher needs to have command not only of factual knowledge, but also of knowledge about doing and reasoning. Our model therefore covers three knowledge areas: (1) declarative knowledge about facts, rules, or principles, (2) procedural knowledge about how to proceed in particular situations, and (3) conditional knowledge about the reasons

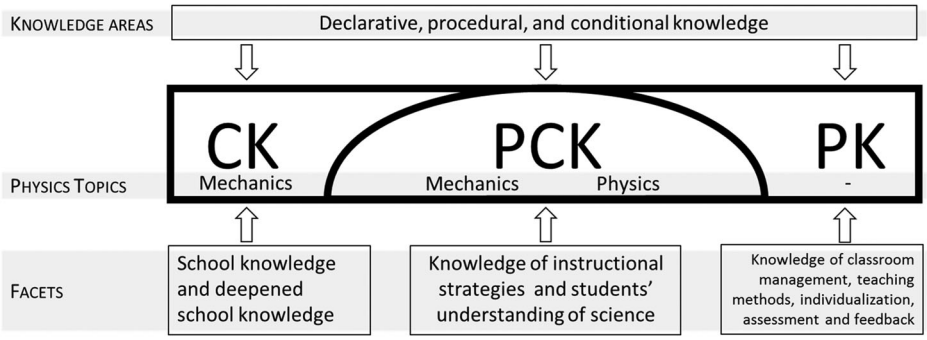


Figure 2. Model of professional knowledge. CK, PCK, and PK are characterized through the same three knowledge areas and distinctive facets. CK and PCK items cover mechanics, PCK items cover also physics in general.

for rules or processes, as well as about judgment of teaching situations (Tepner et al., 2012; cf. Paris, Lipson, & Wixson, 1983). Table 2 shows examples for CK and PCK for all three knowledge areas.

Facets

As already discussed in this paper, PCK is further defined by its facets which may vary between different models. Because the facets *knowledge about student understanding* and *knowledge about instructional strategies and representations* are widely regarded as important facets of PCK, we included these in our model (Tepner et al., 2012). The facet knowledge about student understanding includes knowledge about students' preconceptions, their reasons, and instructional strategies to engender students' understanding. The facet knowledge about instructional strategies and representations includes knowledge about experiments which are, or at least should be, a major component of science lessons (Lunetta, 1998).

CK and PK are also characterized by their facets. For CK, science teachers should have command of CK at the school level. However, there are different ideas about the level of

Table 2. Examples for PCK and CK tasks of different knowledge areas.

Knowledge area	Examples
Declarative knowledge	**What** • are students' alternative conceptions concerning force? • is the equation for the lever principle?
Procedural knowledge	**How** can you change the experiment in a way that • students can identify the relationship between the two variables? • a researcher gets a valid result?
Conditional knowledge	**Why** • is it important to consider students' preconceptions in lessons? • does the summer last longer on the northern hemisphere than on the southern hemisphere?

CK needed for teachers to teach a subject (Baumert et al., 2010). Our model covers the CK facets *school knowledge* and *deepened school knowledge*. School knowledge stands for physics knowledge taught to students aged 12–18. For some teachers, this exceeds the school knowledge they actually teach. Deepened school knowledge describes a deeper understanding of physics concepts taught at school (Kind, 2014) which is essential for preparing physics lessons and integrating it in the structure of a teaching unit and the discipline. Deepened school knowledge is a kind of CK especially for teachers which should qualify them to design tasks for students and to answer questions asked by their students that are beyond the current topic; deepened school knowledge does not rely on higher mathematics (Kirschner, 2013; Woitkowski, Riese, & Reinhold, 2011). Because we assume greater impact of school knowledge and deepened school knowledge on middle school teaching than that of university level knowledge on such teaching, the university level knowledge was not considered as a facet in our model.

For PK, 'definitions and specifications' vary (Voss, Kunter, & Baumert, 2011, p. 953). Grossman (1990) summarized that PK includes knowledge and beliefs about learning and learners in general as well as about 'aims and purposes of education' and 'knowledge of principles of instruction' (p. 6). Accordingly, PK is further composed of knowledge and skills related to classroom management. Voss et al. (2011, p. 953) defined the following facets in an attempt 'to find a unifying conceptualization': 'knowledge of classroom management, knowledge of teaching methods, knowledge of classroom assessment, knowledge of learning processes, and knowledge of individual student characteristics'. For PK, knowledge about *classroom management*, *teaching methods*, *individualization*, and *assessment/feedback* are facets of our model (Lenske, Thillmann, Leutner, & Wirth, 2015), as these seem to be core components of PK.

Topics

Similar to Veal and MaKinster (1999), we distinguish between domains (like physics or biology) and topics (like mechanics or electricity). CK as well as some aspects of PCK are considered as topic-specific, for instance, student understanding of particular physics ideas or instruction for teaching these ideas. PCK comprises knowledge about how to teach particular topics and how particular topics are learned (Gess-Newsome, 2015, p. 36). The topic chosen for our model is mainly about mechanics because it is one major part of the German university and school physics curriculum at different age levels; and we expected that all teachers had experiences with learning and teaching mechanics. From a physics perspective, concepts of mechanics, particularly those about forces and how these relate to changes in velocity, are relevant to other physics topics. For instance, forces are relevant for thermodynamics (e.g. pressure from particles interacting with container walls), and for electricity (forces between charges). In our test instrument, CK and PCK questions were constructed on the same topic since we hypothesized a closer relationship between CK and PCK.

In addition to being specific for particular topics, PCK also includes components that are domain-related but cover more than a specific topic such as domain-general knowledge about reasons for conducting experiments in physics lessons. This approach is taken into account in the model.

In contrast to PCK and CK, PK is assumed to be subject- and topic-independent (Voss et al., 2011).

Methods

The model described above (see Figure 2) was used to construct a paper-and-penciltest instrument containing three tests (PCK test, CK test, and PK test) for physics teachers' professional knowledge. The model was our frame of reference for the selection and development of tasks to assess professional knowledge. To evaluate the tests, three consecutive studies (a pilot, the first study, and the second study) were conducted. For the analysis reported in this paper, the data of the second study were investigated with Rasch analysis. This section describes the development of the PCK test and explains how its validity was evaluated. Descriptions of the considerations and methodical decisions we made might assist other researchers in using the instrument, adapting it to their needs, or in constructing new instruments.

Instrument

The PCK test (Kirschner, Borowski, & Fischer, 2011a) consists of 17 items, most of which are open-ended, and will be described further in this section. Figure 3 presents three examples of the PCK items. Table A1 in the Appendix presents a list of all the PCK items and their characteristics concerning the model. All the PCK items and the coding schemes for the example items of Figure 3 can be found in Table A2. The CK test (Kirschner, Borowski, & Fischer, 2011b) consists of 15 multiple-choice and open-ended items which were also coded with a coding scheme. The reliability and separation of the CK test ($\mathrm{Rel}_{\mathrm{CK}}^{\mathrm{person}} = .76$, $\mathrm{Rel}_{\mathrm{CK}}^{\mathrm{item}} = .98$, $\mathrm{Sep}_{\mathrm{CK}}^{\mathrm{person}} = 1.80$, $\mathrm{Sep}_{\mathrm{CK}}^{\mathrm{item}} = 7.33$) are appropriate. For declarative PK ($\mathrm{PK_d}$), nine complex-multiple-choice questions were used; for conditional and procedural PK ($\mathrm{PK_{cp}}$), different courses of action for 14 teaching situations have to be rated by the teachers (Lenske et al., 2015; Thillmann, 2010). Because the person reliability of the $\mathrm{PK_d}$ test is not satisfactory, classical test scores were used (Cronbach's $\alpha_{\mathrm{PK\_d}} = .90$). The reliability and separation of the $\mathrm{PK_{cp}}$ test are good ($\mathrm{Rel}_{\mathrm{PK\_cp}}^{\mathrm{person}} = .72$, $\mathrm{Rel}_{\mathrm{PK\_cp}}^{\mathrm{item}} = .96$, $\mathrm{Sep}_{\mathrm{PK\_cp}}^{\mathrm{person}} = 2.45$, $\mathrm{Sep}_{\mathrm{PK\_cp}}^{\mathrm{item}} = 5.03$). In our study, the CK and PK tests were only used for validating the PCK test.

From the model to the PCK test

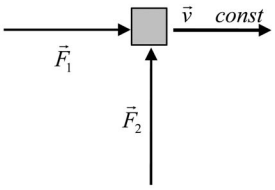
In accordance with the model, the PCK test covers tasks and items that are more related to the subject physics (more toward the left-hand side of PCK in Figure 1), whereas other items are more general (more toward the right-hand side of PCK in Figure 1). Item 1 in Figure 3 is an example that is more tightly related to CK because the correctly drawn force vector is not given. The item may have been more balanced if a correct student answer was presented. Item 10 is a more balanced PCK item as neither content nor pedagogy predominates.

Item 10 and Item 13 in Figure 3 were assigned clearly to one *knowledge area*. It was not always possible to assign each item to only one knowledge area because some items ask for more than facts but not explicitly for processes or reasons (e.g. Item 1 in Figure 3). Nevertheless, we ensured that all knowledge areas – declarative, procedural, and conditional – were covered. Roughly two-thirds of the PCK items cover the *topic* mechanics (e.g. Item 1 and Item 10 in Figure 3). Restricting the test to a few subtopics was needed in order to limit test length. The chosen subtopics in mechanics were velocity/speed and

Item 1 (Knowledge area: Conditional knowledge, Topic: Physics, Facet: Knowledge of instructional strategies):
Why do you use experiments in physics lessons? Please give as many reasons as possible.

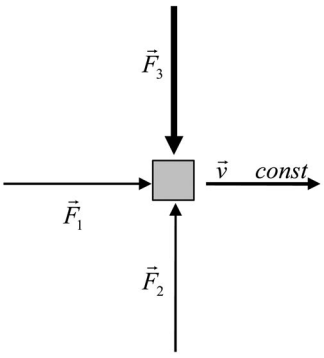
Item 10 (Knowledge area: Declarative knowledge, Topic: Mechanics, Facet: Knowledge of students' understanding of science):
Students may have misconceptions about the concepts *speed* and *velocity*. Write down as many misconceptions as possible.

Item 13 (Knowledge area: Not assigned, Topic: Mechanics, Facet: Knowledge of student's understanding of science):
A colleague has presented the following test item:



Consider an object that is moving without friction at constant velocity in straight line.
As the object moves, it is under the influence of 3 forces.
$\vec{F}_1$ and $\vec{F}_2$ are given. Draw the third force vector, $\vec{F}_3$.

A student has submitted the following drawing:



What misconceptions are evident in the student's answer?

Figure 3. Example items to assess PCK. The coding schemes can be found in the Appendix.

forces (also for circular motion), as well as the relationship between forces, energy, and power. These topics refer to typical student learning difficulties (understanding Newton's Laws, disentangling force, energy, and power) that teachers should recognize and for which they should have ideas for an appropriate instructional approach. Also, it was likely that physicists might have come across learning difficulties in mechanics. When validating the PCK test, these subtopics could help to identify whether knowledge

about student learning difficulties and corresponding instruction is specific for teachers or any person who is trained in physics (see Known Groups Construct Validity). In order to have the opportunity to investigate the relationship between this test and another German PCK test concerning another topic (Olszewski, 2010), two items on electricity were included. Another third of the PCK items asks questions related to physics in general (Item 13 in Figure 3 as an example). Most items are assigned to one *facet* (see examples in Figure 3) but as PCK can be seen as the integration of facets (Park & Chen, 2012), single items can cover more than one facet (compare them in Table A1).

The model makes no predictions about item difficulty. We do not expect that items concerning knowledge about students' understanding are easier or harder to solve than items about instructional strategies. Although different aspects of mechanics may vary in difficulty, the broad categories 'mechanics' versus 'physics' are not expected to influence item difficulty in a particular way. It is difficult to predict possible influence of the knowledge area. In order to investigate the influence of any variable (e.g. knowledge area) on item difficulty, it is necessary to create many items which vary with this variable while keeping constant as many other variables as possible (e.g. subtopic, facet, length of item description, figures). As the items had not been developed in this structure and the number of items was limited, it was not possible for us to test the influence of the model's components on item difficulty.

Test format, coding, and test taking

Closed- and open-ended test formats are used in PCK tests (e.g. Dollny, 2011; Schmelzing et al., 2013). In order to collect as much information as possible with a paper-and-pencil test, we decided to use mainly open-ended questions for testing PCK, even though an objective analysis might create a challenge and requires a sound coding system.

In German language PCK tests, it is common to utilize vignettes that contain a realistic description of a teaching sequence or teaching material (e.g. Brovelli, Bölsterli, Rehm, & Wilhelm, 2013; Riese, 2009a). Our PCK test includes five vignettes that mainly require open-ended answers (see example Item 1 in Figure 3). This type of items can refer to the current definition of PCK – 'Knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes' (Gess-Newsome, 2015, p. 36).

All the open-ended items were coded by two raters using a manual. For every item, a coding scheme was developed in order to identify correct, partly correct, or incorrect answers (see Appendix, Table A2, for the example items of Figure 3). Partial credit was given for knowledge of part of the solution. Partial credit was considered in all example items. Coding options developed for the open-ended items were derived from the literature where possible. Not for all apparently correct answers could theoretical or empirical evidence be found. For some items, it was difficult to decide if some answers are correct or incorrect. Whereas we ended up with satisfying coding options for some items (see Appendix, Table A2, Item 1), we had difficulties with other items (see Appendix, Table A2, Item 13). In order to enhance the inter-rater agreement, the coding manual lists many examples for every correct answer that were partly derived from teacher responses as they appeared in the pilot and the first study.

The entire test instrument was completed voluntarily by the participants in supervised sessions lasting two hours. The duration of the PCK test was limited to 40 minutes. It was

intended but not guaranteed that most participants would complete all the items. The first three items of the PCK test were to be completed in one minute each to highlight the speed test nature of the test.

Content validity

Content validity is 'the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose' (Haynes, Richard, & Kubany, 1995, p. 238). Content validity needs to be evaluated before a final version of a test is created, whereas other aspects of validity, such as construct validity, can be analyzed with data from implementation of the final version.

In the study reported in this paper, content validity was considered in two steps. First, a model is required from which the test itself is then developed. Basically, the paper so far has described this first step in which the model took aspects into account that are widely accepted and the test was then constructed based on the model. Second, each PCK item from the test was checked to determine if its content is (or at least may be) highly relevant to physics lesson planning or teaching. For this second step, the items and sometimes the related coding scheme were discussed by experts (six university physics educators[1]and three experienced physics teachers) focusing on the question, 'Should physics teachers know the content of this item?' We assume a reasonable extent of content validity because both steps were conducted convincingly.

Known groups construct validity and sample

The further validation of the PCK test was guided by the question 'Does the PCK test measure knowledge that physics teachers, but not other groups of participants have (Thorndike & Thorndike-Christ, 2014)?' In order to answer this question, four groups of participants were tested: teachers who taught different school subjects (physics vs. biology, chemistry, or math), teachers from different school types (basic education teachers vs. higher education teachers[2]), pre-service teachers at a university, and physicists (see description of the sample in Table 3). In order to further validate the test, assumptions were made about differences in PCK that can be expected between the groups. The following differences were assumed:

Hypothesis 1: Physics teachers perform better on a physics-specific PCK test than do other science or math teachers who are trained for the same school track with a similar duration of study at university but without preparation in physics.

Hypothesis 2: Higher education physics teachers (HE) who are allowed to teach lower and upper secondary perform better on a physics-specific PCK test than do basic education physics teachers (BE) who only teach lower secondary (cf. Bonsen, Bos, & Frey, 2008 for the distinction). This hypothesis seems to be plausible because students' as well as pre-service and in-service teachers' achievement varies depending on their school track, with students who learn and teachers who teach at higher level school tracks having higher competencies (for students, see Baumert et al., 2010; Jakubowski, 2010; Rönnebeck, Schöps, Prenzel, Mildner, & Hochweber, 2010; for CK and PCK of pre- and in-service teachers for mathematics, chemistry, and physics, see

Table 3. Sample specification (numbers not adding up to 100% are due to missing information).

Group	*N*	Gender (%) ♀	Gender (%) ♂	Age *M*	Age SD
Physics teachers	186	30	71	44	10
Physics teachers HE[a]	149	31	69	44	10
Physics teachers BE[b]	37	22	76	45	10
Non-physics teachers	21	48	52	38	10
Pre-service physics teachers	79	41	57	24	5
Physicists	7	14	86	30	1

Notes: Physics teachers HE/BE are subsamples from Physics teachers.
[a]Physics teachers for higher education.
[b]Physics teachers for basic education.

Baumert et al., 2010; Dollny, 2011; Krauss, Neubrand, Blum, & Baumert, 2008; Riese, 2009a).

Hypothesis 3: In-service physics teachers perform better on a physics-specific PCK test than do pre-service physics teachers at university because the latter group has nearly no teaching experience and has not yet finished their theoretical education.

Hypothesis 4: Higher education physics teachers (HE) perform better on a physics-specific PCK test than do physicists with a similar duration of university study who neither teach at school nor at university (Olszewski, 2010) because the latter group usually has only little contact with school and pedagogy.

Internal structure of professional knowledge and data analysis

CK, PCK, and PK are regarded as distinct dimensions in our model. To explore the details of the internal structure of physics teachers' professional knowledge, a validity check (Messick, 1998) was performed using a multidimensional Rasch analysis working with ACER ConQuest (Wu, Adams, & Wilson, 1998) comparing different models with a varying number of dimensions. In Rasch analysis, a smaller final deviance, Bayes Information Criterion (BIC), and Consistent Akaike Information Criterion (CAIC) indicate a better model fit (Wu et al., 1998; for a discussion of the information criteria see Burnham & Anderson, 2004). If CK, PCK, and PK are distinct dimensions, the correlation between PCK and CK/PK should be higher than that between CK and PK.

The PCK test was investigated with Rasch analysis. Important advantages of Rasch analysis are that the interval scale of item difficulty and that of person ability are established and that items and persons can be mapped on one scale, the item map (Boone & Scantlebury, 2006; for an introduction to Rasch analysis see Boone, Staver, & Yale, 2014). Fit statistics and person measures for PCK were calculated with Facets[3] (Linacre, 2014) taking into account two raters (Boone, Townsend, & Staver, 2016); missing data were treated as missings (Linacre, 2015, p. 647). Three items failed in one criterion (model fit or objectivity) and, therefore, were not further analyzed (see Table 4 for an overview about item statistics). The received person measures were standardized ($M = 2.00$, $SD = 0.75$ for the whole sample) and further analyzed with *t*-tests, Pearson's correlation, and Cohen's *d*. Normal distribution is assumed for either $N > 30$ or an adequate value of the Kolmogorov Smirnov test.

Table 4. Item measures and fit values.

Item		Measure	Error (model S.E.)	Outfit MnSq	Outfit ZStd
1		−0.84	0.09	0.54	−8.83
2		0.92	0.10	1.06	0.80
3	Excluded from the analysis because of missing objectivity ($ICC_{2,2}$(unjust) < .7)				
4		0.10	0.09	0.80	−3.21
5		−0.36	0.09	1.22	3.20
6		0.94	0.10	1.11	1.46
7a		−0.20	0.09	1.03	0.43
7b		0.10	0.09	1.27	3.85
8a		−1.10	0.09	0.84	−2.64
8b		−0.17	0.09	1.00	0.04
9	Excluded from the analysis because of fit criteria	−1.53	0.09	1.62	8.20
10		−0.47	0.10	1.10	1.41
11		0.61	0.10	0.94	−0.91
12	Excluded from the analysis because of missing objectivity ($ICC_{2,2}$(unjust) < .7)				
13		0.75	0.10	0.87	−1.91
14		−0.30	0.10	0.94	−0.82
15		0.00	0.09	1.18	2.59

Notes: Measures express the item difficulty; low values indicate easy items, high values correspond with hard items. MnSq: Mean-Square fit statistics is the measure for the fit between model and data. ZStd: Z-Standardized reports the significance of the (mis)fit. Fit criteria are 0.5 < MnSq < 1.5 (item is productive for measurement) or ZStd < 2.0 (a potential item misfit is not significant) as recommended by Linacre (2015, p. 602).

Results

Objectivity

All the items analyzed had satisfactory inter-rater agreement based on the calculations from ten percent of the sample coded by two trained raters. The values of Goodman's and Kruskal's Gamma and $ICC_{2,2}$(unadjusted) were higher than .7.

Dimensions of professional knowledge, reliability, and item map

The results of the multidimensional Rasch analysis are shown in Table 5. It revealed that CK, PCK, and PK_d and PK_{cp} were distinct dimensions because the final deviance was significantly smaller for the four-dimensional model than for the other models (Chi-square tests, $p < .001$), and BIC and CAIC results further supported this four-dimensional model. The correlation between the dimensions of CK and PCK ($r = .521$, $p < .001$) was significantly higher than that between CK and PK_d ($r = .202$, $p < .01$) and CK and PK_{pc} (r

Table 5. Degrees of freedom and final deviance for possible models with all items, calculated with in-service physics teachers.

Model	DIM 1: CK and PCK and PK	DIM 1: CK and PCK DIM 2: PK	DIM 1: CK DIM 2: PCK and PK	DIM 1: CK DIM 2: PCK DIM 3: PK	DIM 1: CK DIM 2: PCK DIM 3: PK_d DIM 4: PK_{cp}
f	435	437	437	440	444
dev	130197	129368	129307	129108	128685
BIC	132172	131352	131291	131105	130700
CAIC	131619	130797	130736	130547	130137

Notes: f = degrees of freedom; dev = final deviance; DIM 1–4 = dimension 1–4.

= .039, $p = .596$). The correlation between PCK and PK_{cp} ($r = .232$, $p < .01$) was significantly higher than that between CK and PK_{cp}, and the correlation between PCK and PK_d ($r = .239$, $p = .001$) was of similar strength as that between CK and PK_d.

The person reliability, as a measure for the internal consistency, was satisfactory ($Rel_{PCK}^{person} = .76$). The item reliability was good ($Rel_{PCK}^{item} = .98$). The person separation was alright with $Sep_{PCK}^{person} = 1.78$ and the item separation was good with $Sep_{PCK}^{item} = 6.46$. Figure 4 shows the item map with the plots of all the physics teachers' ability. On the left-hand side are the plots for the teachers that show their increasing ability. On the right-hand side are the plots for the items that show their increasing difficulty. On average, a teacher can solve an item that is on the common scale (item difficulty and teacher ability) on the same level more likely compared to solving other items whose difficulty is above the teachers' ability. In order to identify teachers with different ability, items with varying difficulty are needed. The item difficulty shown in Figure 4 matched person ability over the whole range with some gaps between items and three items with the same difficulty. As expected, the model of physics teachers' professional knowledge did not predict item difficulty (ordering or spacing of the items on the map).

Comparison of known groups

The PCK measures of the different groups (see Table 3) can be found in Figure 5. Comparison 1 shows the mean PCK measures of physics teachers for higher education contrasted with the mean PCK measures of non-physics teachers. Additionally, the standard error (+/− 2 *SE*) is indicated with error bars. Likewise, comparisons 2–4 show the comparison of mean PCK measures of physics teachers for higher and basic education (comparison 2), of in-service and pre-service physics teachers (comparison 3), and of physics teachers for higher education and physicists (comparison 4). Hypothesis 1 was confirmed (see Figure 5, comparison 1) because as expected, non-physics teachers achieved lower values than did physics teachers, $t(16) = 5.85$, $p < .001$, $d = 1.4$. Hypothesis 2 was also confirmed (see Figure 5, comparison 2) because teachers at lower level school tracks and with less university content preparation received lower values, $t(183) = 7.21$, $p < .001$, $d = 1.3$. Hypothesis 3 was confirmed as well (see Figure 5, comparison 3) because in-service physics teachers achieved better results than did pre-service physics teachers, $t(261) = 3.86$, $p < .001$, $d = 0.5$. Hypothesis 4 was confirmed, too (see Figure 5, comparison 4) because the PCK of the seven physicists who neither teach at school nor at university scored lower than did physics teachers at higher level school tracks, $t(154) = 2.53$, $p < .05$, $d = 1.0$.

Discussion

In this paper, we focus on the development of a model of physics teachers' professional knowledge, describe the development of a corresponding test instrument to confirm the dimensionality of the chosen construct of professional knowledge (CK, PCK, and PK), and include evidence for its validity. We also describe the model, its operationalization, the design of the study, the results, and some crucial points that make it difficult to assess PCK. In the following discussion, the results of each step are evaluated and implications for future research considered.

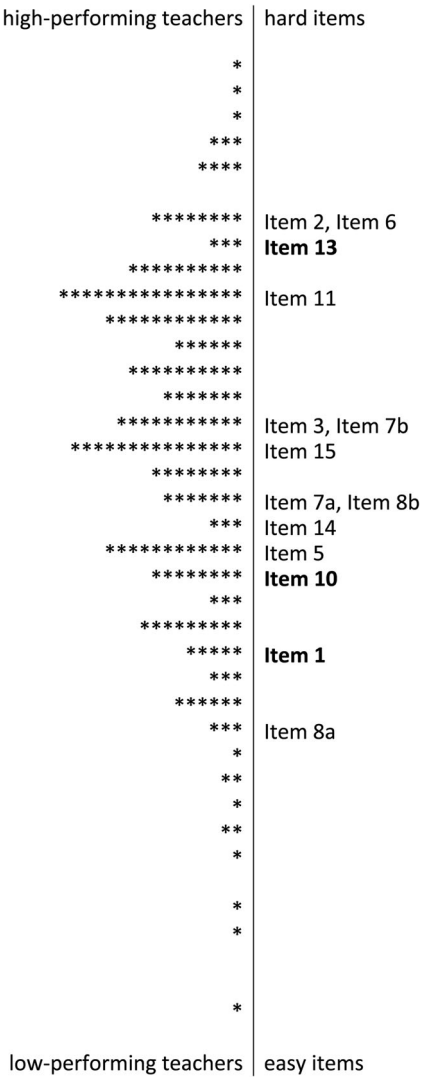


Figure 4. Item map. On the left side the teacher ability is shown. Every * indicates one teacher. The item difficulty is shown on the right side. Higher person measures and higher item difficulties are presented at the top of the map. Example items are printed bold.

Paper-and-pencil tests. These tests promise generalizable results at the expense of measuring only verbalized, canonical knowledge on action. We hope to identify the knowledge teachers need by analyzing their knowledge, their instructional strategies, and student outcomes in large samples. We recognize the risk that test scores alone are over-interpreted and misunderstood as a measure for good teaching. We hope that such tests will not be used in the assessment of teaching or result in 'teaching to the test' in teacher education.

Modeling the relationships between CK, PCK, and PK theoretically. CK, PCK, and PK are assumed to be distinct dimensions of professional knowledge, whereas the connection between PCK and CK/PK is tighter than that between CK and PK (see Figure 1). PCK

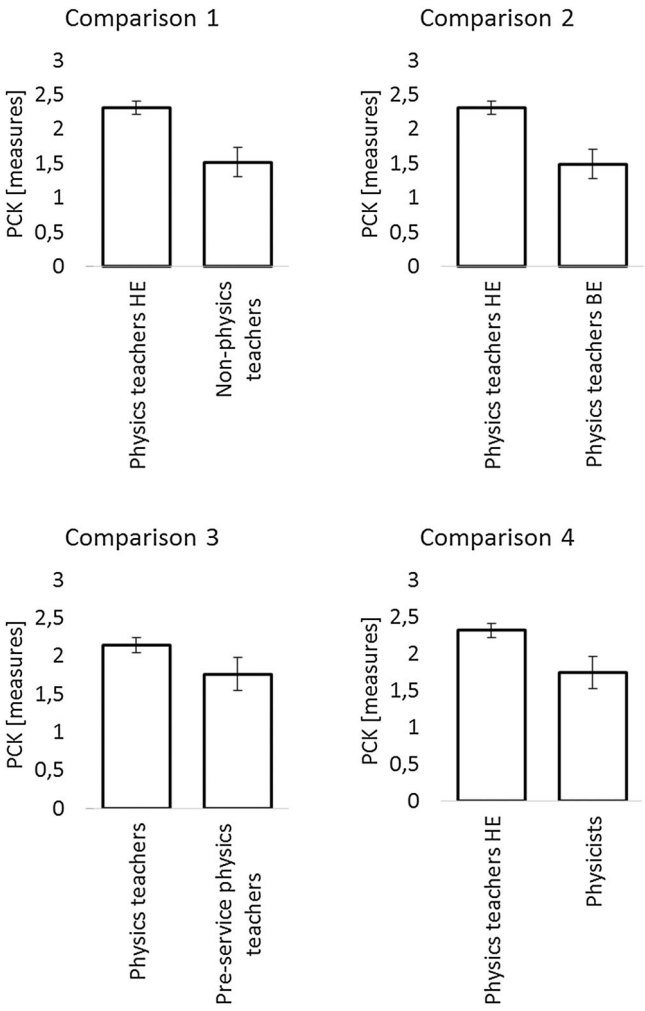


Figure 5. Performance in the PCK Test. Comparisons 1–4 show the PCK measures of different groups. Error bars indicate +/− 2 *SE*. Physics teachers HE (Physics teachers for higher education) and physics teachers BE (Physics teachers for basic education) are subsamples from physics teachers.

tasks and items can be designed along a continuum of being more subject-related to more general, but the items in this study were not assigned along such continuum and were simply included in the dimension PCK. In future research, it would be interesting to ask experts to more carefully classify PCK items on this continuum and analyze the correlation of CK/PK with separate groups of PCK items with different coverage of CK/PK aspects.

Operationalization of CK, PCK, and PK. A focus of this paper is on the model (see Figure 2) and how the model acted as a framework for the development of the tests. The model was used as a guideline to ensure content validity and to make underlying assumptions transparent to other researchers. As discussed above, our model does not cover all potential aspects of PCK that are considered in the literature. As a result, our model and the PCK test are limited to those aspects of PCK that we assume to be

related to PCK as a whole. This issue applies to all measures of PCK. Another limitation has to do with the limited number of test items. The item map suggested that the item difficulty in our test was not ideal; the difficult items should be distributed more equally. This limitation can be addressed by giving more or less information in the item stem and adapting the rating scales. The choice of only one topic and a few subtopics is a basic limitation of our model because we cannot make any statements other than on those components tested in this study. Although we assume that knowledge about learning and teaching of one topic helps to establish knowledge of learning and teaching in another topic, little is known about whether and how teachers use knowledge across various topics. A first sensible expansion of the model is to include more subtopics (e.g. momentum). A second one is to include another physics topic and to create additional tests using the expanded model. A second test and data set would allow researchers to additionally test for convergent validity because a test counts as valid if the results of different PCK tests correlate substantially (Kirschner, Taylor, Rollnick, Borowski, & Mavhunga, 2015; Olszewski, 2010). A comprehensive test would not only cover diverse facets and topics but would also use more elaborated vignettes with a higher degree of complexity. Vignettes could be presented in the form of text or video which would ideally refer to the teacher's own teaching. In order to have the opportunity to adapt the tests to teachers' actual experiences, a set of similar items for different topics and age levels, from which researchers can choose, would be desirable. Such a set could also be used to control for the effect of familiarity that could confound the results when the only items used are about a topic that a teachers teaches in the current school year or about a topic that a teacher has never taught before.

Content validity was ensured by the model-based development of the test instrument and the consultation of experts. The definition of the knowledge areas of the model was partly fuzzy, so it was not possible to assign every item to all aspects of the model.

Structural validity was established. It was confirmed by a multidimensional Rasch analysis that CK, PCK, and PK are distinct dimensions. All dimensions showed a satisfactory reliability. The conclusion about validity was supported by the correlations between the dimensions. Overall the correlations between PCK and CK/PK were higher than between CK and PK.

Construct validation was guided by testing the differences in measures of teacher knowledge across in-service physics teachers and other groups. The results of the hypotheses tests were as expected and support the assumption of a valid test: (i) on average, non-physics teachers scored lower than did physics teachers, (ii) teachers who teach at lower level school tracks scored lower than did physics teachers at higher level school tracks, (iii) physics teachers scored higher than did pre-service physics teachers, and (iv) seven physicists, who do not teach at school and have no teaching background, scored lower than did the teachers. The last result is promising but should not be over-interpreted because of the small sample size in our study. Unfortunately, it is difficult to obtain a sample of physicists who have no connections to university or schools as participants.

Validity is not dichotomous, that is, not just about the test being valid or non-valid (cf. Messick, 1989). We attempted to ensure that the PCK test does not measure cognitive abilities, CK, PK, computational skills, or science knowledge, but only knowledge that distinguishes physics teachers from other groups of participants. However, the major aim of studying PCK is the expectation that higher PCK leads to better student achievement

(Abell, 2008). In order to claim anything about the practical usefulness of the test, predictive validity must be evaluated. As we were not able to analyze the interaction between teachers' PCK and student outcomes, we do not draw any conclusions for school and teacher education (cf. Messick, 1989).

The findings of our study demonstrate that physics teachers' CK, PCK, and PK are distinct dimensions of teacher professional knowledge, which implies that we can continue to use this distinction in further studies. This distinction is especially important because it demonstrates that PCK can be seen as an independent construct. From a methodical point of view, we have described a normatively set model and the related test development to offer options to other researchers. Because we assume that teachers' professional knowledge influences teaching and student outcomes, our hope is to identify ways to best support physics teachers and improve the quality of their instruction. One way to do this is to identify the knowledge that teachers need and use that to influence the content taught at teacher education programs. This includes the analysis of the professional knowledge currently taught at teacher education programs. We still need to gain more insight into what and how professional knowledge impacts teaching. To close the research gap between PCK and its relevance for teaching, we suggest further studies for analyzing the predictive validity of PCK measures on student outcomes.

Notes

1. All physics educators own a PhD in physics education, three of whom were professors for physics education at the time, now all of them are.
2. As teacher education in Germany differs between federal states and changes frequently, the number and quality of physics courses, physics education courses, and pedagogy courses all vary. This variation is making the characterization of the teachers recruited into our study imprecise. We, therefore, only distinguish between basic education physics teachers (BE) who only teach lower secondary, and higher education physics teachers (HE) allowed to teach lower and upper secondary. All pre-service teachers complete a practical pedagogical training after their university degree studies.
3. Unfortunately, Facets does not allow a principal component analysis of residuals for further analysis of the dimensionality.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Bundesministerium für Bildung und Forschung.

Notes on contributors

Sophie Kirschner is a postdoctoral researcher at the Justus Liebig University Giessen. Her main interest is teacher knowledge and learning with a particular focus on physics teachers. She received her diploma in physics in 2009 at the Goethe University Frankfurt. In 2013 she earned her Ph.D. at the University of Duisburg-Essen with the dissertation Modeling and Analysing Physics Teachers' Professional Knowledge.

Andreas Borowski is professor for physics education at the University of Potsdam, Germany. His research interests are the professional knowledge of (pre-service) teachers and the competence of physics students at upper secondary and university.

Hans Ernst Fischer is Senior-Professor for physics education at the Faculty of Physics at the University of Duisburg-Essen. His main research focus is on teaching and learning physics from primary to university level and professional knowledge and professional development of pre- and in-service teachers.

Julie Gess-Newsome is the Interim Academic Dean at Oregon State University-Cascades in Bend, Oregon, USA. Her research interests include teacher professional knowledge involving pedagogical content knowledge.

Claudia von Aufschnaiter received her PhD in physics education in 1999. She has been working as a full professor of physics education at the University of Giessen since 2007. In her research, she investigates how students develop physics concepts, argue about physics, and experience working on physics tasks. Likewise, she investigates how (pre-service) teachers develop "educational concepts" about assessment of student thinking and learning.

References

Abell, S. K. (2007). Research on science teachers' knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah, NJ: Lawrence Erlbaum Associates.

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, *30*(10), 1405–1416.

Alonzo, A. C., & Kim, J. (2015). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*.

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

Blömeke, S., Felbrich, A., Müller, C., Kaiser, G., & Lehmann, R. (2008). Effectiveness of teacher education. *ZDM*, *40*(5), 719–734.

Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer.: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare* [Pre-service teachers' professional competence: Knowledge, beliefs and learning oportunities of German pre-service teachers at university and school]. Münster: Waxmann.

Bonsen, M., Bos, W., & Frey, K. A. (2008). Germany. In Ina V. S. Mullis (Ed.), *TIMSS 2007 encyclopedia. A guide to mathematics and science education around the world*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center; Lynch School of Education, Boston College.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253–269.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. New York: Springer.

Boone, W. J., Townsend, J. S., & Staver, J. R. (2016). Utilizing multifaceted Rasch measurement through FACETS to evaluate science education data sets composed of judges, respondents, and rating scale items: An exemplar utilizing the elementary science teaching analysis matrix instrument. *Science Education*, *100*(2), 221–238.

Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., … Sumfleth, E. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN) – Kurzdarstellung des BMBF-Projekts [Professional knowledge of science teachers (ProwiN) – A brief outline of the BMBF project]. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 341–349.

Bromme, R. (1994). Beyond subject matter: A psychological topology of teachers' professional knowledge. In R. Biehler, R. W. Scholz, R. Sträßer, & B. Winkelmann (Eds.), *Mathematics didactics as a scientific discipline: The state of the art* (pp. 73–88). Dordrecht: Kluwer.

Brovelli, D., Bölsterli, K., Rehm, M., & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettentest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat [Assessing professional competencies for science teaching: A vignette test using authentically complex teaching contexts and an open-ended answer format]. *Unterrichtswissenschaft*, *41*(4), 306–329.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, *44*(4), 263–272.

Daehler, K., Heller, J., & Wong, N. (2015). Supporting growth of pedagogical content knowledge in science. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 45–59). London: Routledge Press.

Dollny, S. (2011). *Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften* [Developing and evaluating a test for chemistry teachers' subject specific professional knowledge]. Berlin: Logos.

van Driel, J., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*(6), 673–695.

Fletcher, S. S., & Luft, J. A. (2011). Early career secondary science teachers: A longitudinal study of beliefs in relation to field experiences. *Science Education*, *95*(6), 1124–1146.

Friedrichsen, P. (2015). My PCK research trajectory. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 147–161). London: Routledge Press.

Geddis, A. N. (1993). Transforming subject-matter knowledge: the role of pedagogical content knowledge in learning to reflect on teaching. *International Journal of Science Education*, *15*(6), 673–683.

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 3–17). Dordrecht: Kluwer.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). London: Routledge Press.

Gess-Newsome, J., Carlson, J., Gardner, A., & Taylor, J. (2010). Impact of educative materials and professional development on teachers' professional knowledge, practice, and student achievement. Retrieved from http://bscs.org/primepapers

Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. In *IERI Monograph Series – Issues and Methodologies in Large-Scale Assessments* (Vol. 3). Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_03_Chapter_6.pdf

Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik: Modellierung und Testkonstruktion* [Pedagogical content knowledge of pre-service physics teachers at university]. Berlin: Logos.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education. Professional development and practice series*. New York, NY: Teachers College Press.

Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching*, *11*(3), 273–292.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141.

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

Jakubowski, M. (2010). Institutional tracking and achievement growth: Exploring difference-in-differences approach to PIRLS, TIMSS, and PISA Data. In J. Dronkers (Ed.), *Quality and inequality of education* (pp. 41–81). Dordrecht: Springer.

Jüttner, M., & Neuhaus, B. J. (2012). Development of items for a pedagogical content knowledge test based on empirical analysis of pupils' errors. *International Journal of Science Education*, *34*(7), 1125–1143.

Kind, V. (2014). Science teachers' content knowledge. In H. Venkat, M. Rollnick, J. Loughran, & M. Askew (Eds.), *Exploring mathematics and science teachers' knowledge* (pp. 15–29). Abingdon: Routledge.

Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften [Modeling and analysing physics teachers' professional knowledge]*. Berlin: Logos.

Kirschner, S., Borowski, A., & Fischer, H. E. (2011a). ProwiN-Test zum fachdidaktischen Wissen von Physiklehrkräften [ProwiN PCK test for physics teachers]. In A. Borowski, H. E. Fischer, M. Jüttner, S. Kirschner, B. J. Neuhaus, E. Sumfleth, … S. Witner (Eds.), *ProwiN-Testinstrumente*. Essen: Universität Duisburg-Essen.

Kirschner, S., Borowski, A., & Fischer, H. E. (2011b). ProwiN-Test zum Fachwissen von Physiklehrkräften [ProwiN CK test for physics teachers]. In A. Borowski, H. E. Fischer, M. Jüttner, S. Kirschner, B. J. Neuhaus, E. Sumfleth, … S. Witner (Eds.), *ProwiN-Testinstrumente*. Essen: Universität Duisburg-Essen.

Kirschner, S., Taylor, J., Rollnick, M., Borowski, A., & Mavhunga, E. (2015). Gathering evidence for the validity of PCK measures: Connecting ideas to analytic approaches. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 229–241). London: Routledge Press.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). *The professional knowledge of German secondary mathematics teachers: Investigations in the context of the COACTIV project*. Retrieved from http://tsg.icme11.org/document/get/397

Lenske, G., Thillmann, H., Leutner, D., & Wirth, J. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests [Evaluation of a test to capture teachers' pedagogical-psychological professional knowledge]. *Zeitschrift für Pädagogik*, *18*(2), 225–245.

Linacre, J. M. (2014). *Facets computer program for many-facet Rasch measurement*. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2015). *Winsteps® Rasch measurement computer program user's guide* (3.91.0th ed.). Beaverton, OR: Winsteps.com.

Loughran, J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam: Sense.

Lunetta, V. (1998). The school science laboratory: Historical perspectives and contexts for contemporary teaching. In K. Tobin & B. J. Fraser (Eds.), *International handbook of science education* (pp. 249–264). Dordrecht: Kluwer.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teacher. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Dordrecht: Kluwer.

Marks, R. (1990). Pedagogical content knowledge: From a mathematical case to a modified conception. *Journal of Teacher Education*, *41*(3), 3–11.

McNeill, K. L., González-Howard, M., Katsh-Singer, R., & Loper, S. (2015). Pedagogical content knowledge of argumentation: Using classroom contexts to assess high-quality PCK rather than pseudoargumentation. *Journal of Research in Science Teaching*, *53*(2), 261–290.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5–11.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, *45*, 35–44.
Olszewski, J. (2010). *The impact of physics teachers' pedagogical content knowledge on teacher action and student outcomes*. Berlin: Logos.
Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*, 293–316.
Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, *49*(7), 922–941.
Park, S., & Oliver, S. J. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261–284.
Park, S., & Suh, J. (2015). From portraying toward assessing PCK: Drivers, dilemmas, and directions for future research. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 104–119). London: Routledge Press.
Riese, J. (2009a). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften* [Professional knowledge and professional competence of (pre-service) physics teachers]. Berlin: Logos.
Riese, J. (2009b, October 1). PCK-Test and Codebook (Word document).
Rollnick, M., Bennett, J., Rhemtula, M., Dharsey, N., & Ndlovu, T. (2008). The place of subject matter knowledge in pedagogical content knowledge: A case study of South African teachers teaching the amount of substance and chemical equilibrium. *International Journal of Science Education*, *30*(10), 1365–1387.
Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D., & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis PISA 2009 [Competencies in science from PISA 2006 to PISA 2009]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, … P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 117–198). Münster: Waxmann.
Schmelzing, S., van Driel, Jan H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system'. *International Journal of Science and Mathematics Education*, *11*(6), 1369–1390.
Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.
Shulman, L. S. (1987). Knowledge and teaching of the new reform. *Harvard Educational Review*, *57*, 1–22.
Smith, D. C., & Neale, D. C. (1989). The construction of subject matter knowledge in primary science teaching. *Teaching and Teacher Education*, *5*(1), 1–20.
Smith, S., & Banilower, E. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 89–103). London: Routledge Press.
Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, *4*(2), 99–110.
Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., … Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Item development model for assessing professional knowledge of science teachers]. *Zeitschrift für Didaktik der Naturwissenschaften*, *18*, 7–28.
Thillmann, H. (2010). Test zum pädagogischen Wissen von Lehrkräften, unveröffentlicht [PK test for teachers, unpublished].
Thorndike, R. M., & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education (Pearson new international edition)*. Harlow: Pearson.
Veal, W. R., & MaKinster, J. G. (1999). Pedagogical content knowledge taxonomies. *Electronic Journal of Science Education*, *3*(4).

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

Woitkowski, D., Riese, J., & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte [Modelling content knowledge of prospective physics teachers]. *Zeitschrift für Didaktik der Naturwissenschaften*, *17*, 289–313.

Wu, M., Adams, R., & Wilson, M. (1998). *Acer conquest-generalised item response modelling software*. Melbourne: ACER Press.

Appendix. Overview of PCK items, all PCK items, example items with coding scheme and rating scale

Table A1. List of PCK items and their characteristics concerning the model.

		Model		
Item	Content	Knowledge area	Topic	PCK Facet
1	Reasons for the use of experiments in physics lessons	C	Ph	IS
2	Benefits of emphasizing units (like SI) in physics lessons	C	Ph	SU
3[G]	km/h	D	M	SU
4	Vignette: An experiment about velocity which enhances alternative conceptions	Pr	M	IS + SU
5	General criteria to score students' diagrams	D	Ph	IS
6	Vignette: Scoring of example diagrams	Pr	M	IS
7a	Circular motion: Different trajectories you would expect from your students	D	M	SU
7b	Circular motion: Explanation of the student's rationale for each trajectory	–	M	SU
8a	Brightness of light bulbs in a series circuit: Students' answers	D	*	SU
8b	Brightness of light bulbs: Explanation of the student's rationale for the answers	–	*	SU
9	Consideration of students' preconceptions while planning lessons	C	Ph	SU
10	Identification of students' alternative conceptions about velocity	D	M	SU
11[G]	Alternative conceptions enhanced through the words 'a force is exerted'	D	M	SU
12	Explanation of the relations between force, energy and power with examples	Pr	Ph	IS
13	Vignette: Alternative conceptions about force	–	Ph	SU
14	Vignette: Optimizing a given experiment so that the students are able to clearly identify and understand a relation	Pr	M	IS
15	Vignette: How to proceed with a lesson after a failed experiment	Pr	Ph	IS

Note: The items are not translated but characterized by keywords. [G] Items base on the usage of the German language and cannot be translated meaningfully to English. Shortcuts in the model column: –, not possible to assign; *, beyond the model; **D**eclarative knowledge, **Pr**ocedural knowledge, **C**onditional knowledge in the knowledge areas; **M**echanics, **Ph**ysics in topics; Knowledge of **I**nstructional **S**trategies, Knowledge of **S**tudents' **U**nderstanding of science as facets of PCK.

Table A2. All PCK Items and coding schemas/rating scales for example items.

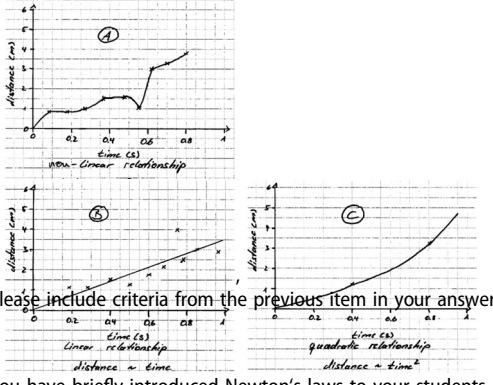
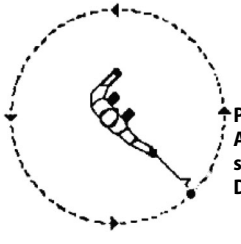
ID	Item
1[S] adapted from Riese (personal communication, 2009b, p. 5)	Why do you use experiments in physics lessons? Please give as many reasons as possible

Example item	*Expectations*
	• Pedagogical/psychological function
	  • Experimentation develops causal and functional thinking and creativity
	  • Experimenting develops the ability to work in an team
	  • Experiments are motivating, increase variety and arouse interest
	  • Experiments make it able to experience learning
	  • Students are actively engaged
	• Subject-related/epistemological function
	  • Experiments support the learning of scientific research methods
	  • Experiments are an established method of gaining knowledge in physics (generating hypotheses and working with them)
	  • Experiments make physical facts visually concrete
	  • Experiments make physical facts/relationships plausible / explain them
	  • Experiments support concept formation
	  • Experiments may lead to cognitive conflicts
	• Practical function:
	  • Students practice handling of data and data analysis
	  • Students practice handling of deviances/establish a relationship to them
	  • Haptic/psychomotoric aspects are developed
	• *Examples of incorrect answers:*
	  • Experiments are required by the curriculum
	  • To practice
	  • To use diverse methods
	• *Rating:*
	  • No correct answer: 0 points
	  • At least one correct answer: 1 point
	  • Correct answers out of two or three categories: 2 points

2[S]	What are the benefits of emphasizing units in physics lessons? Please elaborate.
3[S]	*Items bases on the usage of the German language and cannot be translated meaningfully to English.*
4	A pre-service teacher is planning a lesson on velocity vectors. The pre-service teacher, whom you coach, starts the lesson with an experiment she will perform to introduce the topic of velocity vectors directly after welcoming the students: An electric toy locomotive drives on a straight line.

(*Continued*)

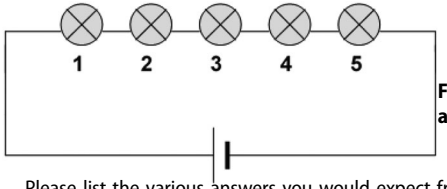
Table A2. Continued.

ID	Item
	Every second, she marks the position of the front of the locomotive. The pre-service teacher asks the students to answer the following written questions silently: What do you notice? What do you conclude? What could the motion be called? What issues can you identify with this lesson, focusing on the formation and deepening of students' preconceptions? Please explain.
5 adapted from Riese (personal communication, 2009b, p. 13)	Imagine you are teaching a lab and students report their results graphically in a diagram using smoothing functions. Please provide a list of general criteria you would use to score students presentation of their results.
6 adapted from Riese (personal communication, 2009b, p. 15)	Imagine that you are planning to teach a lesson whose purpose is for students to experimentally determine that distance is proportional to time squared for an object in free fall. The groups of students present their data in the form of distance-time diagrams and derive the relation with smoothing functions. How do you rate the following three diagrams draw by groups A, B, and C and their conclusions (under the diagrams)?  Please include criteria from the previous item in your answers.
7a, 7b Exercise and figure Hestenes, Wells, & Swackhamer, 1992	You have briefly introduced Newton's laws to your students. In the next lesson, you want them to complete the following exercise:  **Please study the drawing:** **A ball is fixed to a cord that is revolving in a circle. At the given position, the cord breaks. The drawing shows a view from above.** **Draw what the ball's trajectory would be in the picture and explain your rationale.**

Please sketch the various trajectories you would expect from your students. Number the trajectories and then explain the student's rationale for each trajectory based on their prior knowledge.

8a, 8b
(Olszewski, 2010, p. 113)

You would like to discuss the topic, 'Electric current in series and parallel circuits'. The concept of current is already familiar to your students. You will use the following exercise for the lesson:



Five identical lightbulbs are connected to a battery in a circuit. What can you say about the brightness of the lightbulbs?

Please list the various answers you would expect from your students after presenting this exercise. Number their answers so that you can easily refer to them.

What rationales would the students give for their answers? List as many student reasons as you can. The correct physics concept should be listed among them.

9
adapted from Riese (2009b, p. xxiv)

Literature on students learning says that it is important for the learning process to consider students' preconceptions while planning lessons. Please explain why.

10
Example item

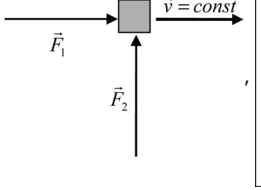
Students may have misconceptions having to do with the physics concepts of *speed* and *velocity*. Write down as many misconceptions as possible.

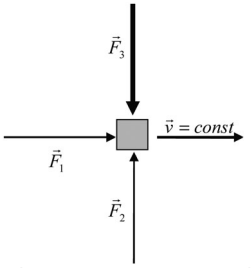
Expectations:

- Misconceptions related to the relationship between distance and time
  - $v = s/t$ always can be used for calculation
  - The formula is $v = s*t$
  - The relationship between v, s, and t is vague [no real misconception]
  - Average speed and mean speed are the same
- Misconceptions related to force
  - A body in motion can cause something / has force; it has more force when it moves faster
  - Without force there is no motion
  - A uniform movement requires a force
  - Bodies become slower by themselves
  - High speed is the result of a large force (neglecting the time aspect)
- Misconceptions related to the direction
  - Velocity and speed are the same
  - Velocity has no direction
  - In a circular motion the velocity is constant
  - Two bodies have the same direction of motion when the have the same goal

(*Continued*)

Table A2. Continued.

ID	Item
	• The direction of force/acceleration and velocity are the same • Negative velocity is not meaningful
	Example of an incorrect answer: • Students have trouble understanding velocity as derivation with respect to time of distance
	Rating: • No correct answer: 0 points • At least one correct answer: 1 point • Correct answers out of two or three categories: 2 points
11	*Items bases on the usage of the German language and cannot be translated meaningfully to English.*
12	Force, Energy and Power are different, although related concepts. How would you explain the difference and relationships between 100 Joules, 100 Newtons, and 100 Watts to your students? Provide meaningful examples and explain the relationships based on these examples. • 100 Joules and 100 Newtons • 100 Newtons and 100 Watts • 100 Watts and 100 Joules
13 Example item	A colleague has presented the following test item:  Consider an object that is moving without friction at constant velocity in straight line. As the object moves, it is under the influence of 3 forces. $\vec{F}_1$ and $\vec{F}_2$ are given. Draw the third force vector, $\vec{F}_3$. A student has submitted the following drawing:



What misconceptions are evident in the student's answer?

Expectations:

- Misconceptions concerning the relationship between force and velocity
  - A force is necessary for every motion
  - Force and velocity/speed are proportional
  - Force and velocity are more or less the same
  - A constant force causes a constant velocity
- Misconceptions concerning the relationship between force and change of velocity
  - A force impacts only the direction of the motion (only velocity not speed)
  - Only the direction of the motion has to be adjusted of the motion
  - Speed does not have to be adjusted with the third force vector

Example of an incorrect answer:

- Students have trouble understanding vector addition

Rating:

- No correct answer: 0 points
- At least one correct answer: 1 point
- Correct answers out of two categories: 2 points

(*Continued*)

Table A2. Continued.

ID	Item
14	A colleague is offering a summer physics course to students in the 14- to 16-year age group. The students have voluntarily signed up for this course. During the week, your colleague carries out a water rocket experiment. In a forest clearing, he and the students use the plastic rockets, a large bucket of water, a yardstick, and an air pump with a pressure gauge in their experiment. The students have notebooks to take notes. The students should be able to recognize the relationship between the amount of water in the rocket and its flight distance. Later, students return to the classroom to determine whether there is a relationship between the amount of water in the rocket and flight distance. Please explain how one could optimize the experiment so that the students are able to clearly identify and understand this relationship.
15 adapted from Riese (2009b, p. xxv)	You would like to introduce a law of physics by conducting a student experiment. After all student groups completed the experiment, there are 20 minutes left before the end of the lesson. The results are so poor that they do not clearly support the law. During the experiment, you had the impression that the students had been working carefully, and you were unable to find any errors. Considering that your goals are to maximize learning opportunities, which of the following tactics would you use to proceed with this lesson? Mark an X next to your choices. • If you have pre-prepared values available, you tell your students that you do not know what they did wrong. You then use the prepared values to tabulate the experiment results. • You tell your students that you cannot work with the results and use modified values. • If the students recognize that their results are poor, you try to find the source of the errors together and apply any recommended changes in a follow-up experiment. • You be honest and tell your students that the experiment did not work as expected, and then you conduct a different experiment. • You postpone the tabulation/analysis of results to the next lesson so that you can think further about it, and decide to start another experiment. • You have the students formulate their own physics law using their current results, and in the next lesson you let them conduct an experiment that proves their formulation wrong. After this, you and your students reflect on all that you have done.

Note: [S] Speed test item.