STEPHAN SCHMELZING, JAN H. VAN DRIEL, MELANIE JÜTTNER, STEFANIE BRANDENBUSCH, ANGELA SANDMANN and BIRGIT J. NEUHAUS

# DEVELOPMENT, EVALUATION, AND VALIDATION OF A PAPER-AND-PENCIL TEST FOR MEASURING TWO COMPONENTS OF BIOLOGY TEACHERS' PEDAGOGICAL CONTENT KNOWLEDGE CONCERNING THE "CARDIOVASCULAR SYSTEM"

Received: 12 March 2012; Accepted: 19 November 2012

ABSTRACT. One main focus of teacher education research concentrates on teachers' pedagogical content knowledge (PCK). It has been shown that teachers' PCK correlates with teaching effectiveness as well as with students' achievement gains. Teachers' PCK should be analyzed as one of the main important components to evaluate professional development programs. On this account, it is necessary to develop standardized measures of biology teachers' topic-specific PCK that are labor-efficient. This paper presents a study on the development, evaluation, and validation of a paper-and-pencil test to measure biology teachers' declarative PCK on the topic of blood and the human cardiovascular system. The development of the test was based, among other considerations, on a review of research literature on PCK and an analysis of 50 videotaped biology lessons. The final test instrument was comprised of 15 items distributed across 2 scales. The findings of the main study—with 93 preservice and in-service biology teachers and 12 biologists—confirmed that this measure of biology teachers' declarative PCK was reliable, objective, and valid. In-service biology teachers scored higher on the test than preservice teachers (effect size Cohen's *d*, 0.65) on one hand and, also, than biologists (Cohen's *d*, 1.00) on the other hand. Future versions of this test should explore enlarging the scales and measuring procedural aspects of PCK.

KEY WORDS: pedagogical content knowledge, teacher assessment, teacher knowledge

Concerns have been raised in both Europe and the USA about whether preservice and in-service teacher education can successfully prepare teachers with the professional knowledge they need to improve teaching practices (Gess-Newsome, Cardenas, Austin, Carlson, Gardner, Stuhlsatz, Wilson et al., 2011). Researchers and policy makers have argued that the enhancement of teachers' professional knowledge leads to high-quality teaching and thereby to gains in student achievement (Baumert, Kunter, Blum, Brunner, Voss, Jordan, Klusmann et al., 2010; Gess-Newsome et al., 2011; Schmidt, Tatto, Bankov, Blömeke, Cedillo, Cogan, Schwille et al., 2007). In particular, educational policy makers in many countries strive to improve teacher education by focusing on teachers' pedagogical content knowledge (PCK). PCK encompasses the idea that teachers have a special notion of content knowledge and general pedagogy, which they draw on in teaching a subject. It describes teachers' understanding of how to help students to understand subject-specific matter (Shulman, 1986, 1987). Therefore, promotion and evaluation of teachers' PCK seem important for improving education systems in the context of global educational competition (Beaton, Martin, Mullis, Gonzalez, Smith & Kelly, 1996; OECD, 2000; Schmidt et al., 2007).

Several recent large-scale studies, in particular, in the domain of mathematics education, have indeed reported positive effects of teacher education programs on teachers' PCK (Schmidt et al., 2007) and positive correlations between teachers' PCK and their instructional quality (Baumert et al., 2010; Hill, Ball, Blunk, Goffney & Rowan, 2007; Park, Jang, Chen & Jung, 2011), as well as between teachers' PCK and students' achievement gains (Baumert et al., 2010; Hill et al., 2007; Hill, Rowan & Ball, 2005; Hill, Loewenberg Ball & Schilling, 2008; Rohaan, Taconis & Jochems, 2009; Staub & Stern, 2002). These studies have provided empirical evidence for a functional relationship between teacher education, teachers' PCK, instructional quality, and students' learning outcomes (cf. Hill et al., 2007).

However, instruments and procedures to explore these relationships in domains other than mathematics education are rare. Therefore, the present study focuses on the domain of science, in particular, on biology education. Currently, there are very few instruments and procedure to measure biology teachers' PCK in a standardized manner that enables generalizable insights (see, however, Gardner & Gess-Newsome, 2011). In particular, labor-efficient instruments and procedures for large sample sizes are needed. Thus, the current study aims to investigate the development, evaluation, and validation of a standardized paper-and-pencil instrument for measuring central components of biology teachers' PCK on the topic of blood and the human cardiovascular system. In particular, we aimed to develop a reliable and objective paper-and-pencil test that has the potential to test biology teachers' declarative PCK in large samples and that measures PCK specifically, distinguishing biology teachers from other biologists.

## Theoretical Framework

### *Conceptualizing Pedagogical Content Knowledge*

PCK describes the unique integration of teachers' content knowledge into their general pedagogical knowledge:

> Within the category of pedagogical content knowledge I include, for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas,

> the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others [ … ] (It) also includes an understanding of what makes the learning of specific concepts easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning. (Shulman, 1986, p. 9).

Shulman (1986) developed a new and trendsetting framework for teacher education that replaced the view of a dichotomous teacher education based on content knowledge and general pedagogical knowledge. Therefore, teacher education programs should also take PCK into account and combine content knowledge and general pedagogical knowledge in order to prepare teachers more effectively. A large number of scholars have worked on the concept of PCK (e.g. Grossman, 1990; Hashweh, 2005; Magnusson, Krajcik & Borko, 1999; Marks, 1990; Park & Oliver, 2008) (see Table 1).

Consequently, there is no consensus about PCK. Attempting to clarify the matter, a series of literature reviews in the late 1990s organized and summarized various PCK conceptualizations (e.g. van Driel, Verloop & de Vos, 1998; Abell, 2008; Gess-Newsome, 1999; Kind, 2009; Lee & Luft, 2008). For example, van Driel et al. (1998) argued that PCK is usually understood as topic-specific knowledge that covers various knowledge components. Nevertheless, different researchers vary as to which components they use for conceptualizing PCK. However, these authors concluded that there is a consensus about two essential knowledge components: (1) knowledge of student learning and conceptions and (2) knowledge of representations and instructional strategies. As shown in Table 1, this conclusion still applies to recent conceptualizations of PCK.

PCK implies two types of knowledge: declarative knowledge and procedural knowledge (Heller, Daehler, Shinohara & Kaskowitz, 2004). Declarative PCK or *knowing that* (Baumert, Blum & Neubrand, 2004; Ryle, 1971) has also been described in terms of *PCK-on-action* (Park & Oliver, 2008), *theoretical–formal PCK* (Fenstermacher, 1994), or *propositional PCK* (Knight, 2002). Declarative PCK is factual knowledge that can easily be expressed in sentences or indicative propositions (Anderson, 1981; Polanyi, 1958). It encompasses propositions, correlations, rules, and theoretical knowledge of ideas and principles and focuses on sense-making and meaning (Knight, 2002). Thus, declarative PCK covers, for instance, factual knowledge of typical students' preconceptions and misconceptions (Baumert et al., 2004).

Procedural PCK or *knowing how* (Baumert et al., 2004) is also described in terms of *craft knowledge* (van Driel et al., 1998), *PCK-in-*

**TABLE 1**

Knowledge components in different conceptualizations of PCK (extended and modified from van Driel et al., 1998)

| *References/PCK components* | *Student learning and conceptions* | *Representations and strategies* | *Goals and purposes* | *Curriculum* | *Evaluation and assessment* | *Media* | *Subject matter* | *Context* | *General pedagogy* |
|---|---|---|---|---|---|---|---|---|---|
| Shulman (1986) | x | x | | | | | | | |
| Tamir (1988) | x | x | | x | x | | | | |
| Smith & Neale (1989) | x | x | x | | | | | | |
| Grossman (1990) | x | x | x | x | | | | | |
| Marks (1990) | x | x | | | | x | x | | |
| Cochran, King & De Ruiter (1993) | x | | | | | | x | x | x |
| Geddis (1993) | x | x | | x | | | | | |
| Fernandez-Balboa & Stiehl (1995) | x | x | x | | | | x | x | |
| van Driel et al. (1998) | x | x | | | | | | | |
| Magnusson et al. (1999) | x | x | x | x | x | | | | |
| Heller et al. (2004) | x | x | | | | | | | |
| Hashweh (2005) | x | x | x | x | x | | x | x | x |
| Loughran et al. (2006) | x | x | x | | | | x | x | x |
| Park & Oliver (2008) | x | x | x | x | x | | | | |
| Lee & Luft (2008) | x | x | x | x | x | x | | | |
| Baumert et al. (2010) | x | x | | x | | | | | |
| Schmidt et al. (2007) | x | x | | x | | | | | |
| Rohaan et al. (2009) | x | x | x | | | | | | |
| Jüttner & Neuhaus (2012) | x | x | | | | | | | |

Table is modified according to the dissertation in German of Schmelzing (2010) (see Acknowledgments) *action* (Park & Oliver, 2008), *skills* (Tamir, 1988), or *practical knowledge* (Fenstermacher, 1994). Procedural PCK, therefore, describes automated skills and action routines that are exercised in the performance of tasks (Anderson, 1981; Polanyi, 1958). Thus, procedural PCK covers teachers' activities during a lesson, for example, if a teacher is able to react appropriately to students' questions and mistakes (Baumert et al., 2004). Contrary to declarative knowledge, procedural knowledge is the ability to do something, which is the reason it is difficult to articulate. It is mainly tacit knowledge that is difficult to transfer to another person by writing it down or verbalizing it (Polanyi, 1958; Stillings, Weisler, Chase, Feinstein, Garfield & Rissland, 1995).

The concept of PCK can be outlined by three cognitive dimensions: (1) *components*, (2) *types*, and (3) *topics* (see Fig. 1; compare also Jüttner & Neuhaus, 2012). The *components* involve knowledge of student learning and conceptions and knowledge of representations and strategies (van Driel et al., 1998); *type* distinguishes between declarative PCK and procedural PCK: between knowledge about what to do and the ability to do it (Anderson, 1981; Baumert et al., 2004). Furthermore, PCK encompasses teachers' ability to adapt subject-related topics and issues to the diverse interests and abilities of their learning group as well as their ability to present subject-related topics for instruction (Shulman, 1986). In this regard, PCK refers to particular *topics*, concepts, problems, and issues. In the present study, the focus is on both components of one type, i.e. declarative PCK, related to the topic of blood and the human cardiovascular system.

### *Measuring Pedagogical Content Knowledge with Paper-and-Pencil Tests*

The complex nature of PCK requires some special and demanding measurement techniques. Therefore, most scholars have concentrated on

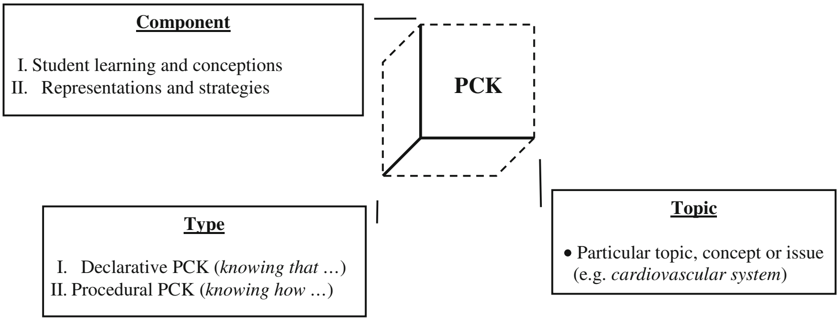

Component
I. Student learning and conceptions
II. Representations and strategies
PCK
Type
I. Declarative PCK (knowing that …)
II. Procedural PCK (knowing how …)
Topic
• Particular topic, concept or issue (e.g. cardiovascular system)

*Figure 1.* Conceptualization of PCK (Schmelzing, 2010; see Acknowledgments)

multimodal approaches to assess teachers' PCK (Baxter & Lederman, 1999). Triangulation of multimodal approaches to PCK often includes structured/semistructured interviews, stimulated recall interviews, (video-)observations, and concept mapping (e.g. Hill et al., 2007; Loughran, Milroy, Berry, Gunstone & Mulhall, 2001; Mulhall, Berry & Loughran, 2003; Park et al., 2011; Piburn & Sawada, 2003). These methods have led to fruitful outcomes and have provided researchers with valid insights into teachers' PCK. Unfortunately, multimodal approaches are only suitable for smaller samples as they are time-consuming and labor-intensive for both participating teachers and researchers. Furthermore, few quality indicators for multimodal evaluations of PCK are available, which makes comparison between various methods difficult (Rohaan et al., 2009). Hence, the challenge in examining teachers' PCK lies in the development of a labor-efficient and time-efficient instrument for larger samples that is also reliable, objective, and valid.

Paper-and-pencil tests are a promising approach for a standardized and quality indicator-oriented method to measure teachers' PCK in large-scale studies. Earlier attempts to evaluate teachers' PCK via paper-and-pencil tests were made by Carlson (1990) and Kromrey & Renfrow (1991). Surprisingly, 20 years later, there are only a few paper-and-pencil tests on teachers' PCK available (Baumert et al., 2010; Hill et al., 2005; Rohaan et al., 2009; Schmidt et al., 2007), especially in the field of biology (Gardner & Gess-Newsome, 2011).

*Item Format.* The use of paper-and-pencil items for measuring teachers' PCK raises several issues, one of which is the adequacy of the item format. For example, the use of closed item formats like multiple-choice items raises problems with developing and evaluating correct answers and distracters. There is no "gold standard" for item format because of multiple normative goals and a lack of empirical evidence (Baxter & Lederman, 1999; Kromrey & Renfrow, 1991). Thus, it is difficult to judge between a "right" and a "wrong" PCK response. Some researchers refer to professional consensus as a template for item scoring (Carlson, 1990; Rohaan et al., 2009). A further problem of closed items is that PCK comprises teachers' individual teaching experiences. A closed item format may run the risk of excluding particular teaching experiences. This means a closed item format covers a limited selection of answer alternatives, but it is not guaranteed that the given selection of answer alternatives will cover teachers' particular teaching experiences or teaching approaches. Moreover, the provision of alternatives in closed item formats could bias responses and help identify the right answer by comparing given alternatives (Hill et al., 2008).

Some of these issues can be avoided by using open-ended items for measuring PCK (Baumert et al., 2010; Heller et al., 2004). Open-ended items, in contrast to closed items, provide an opportunity to measure teachers' individual teaching experiences and exclude response bias problems. Unfortunately, the use of open-ended items for measuring PCK leads to increased work intensity as well as problems associated with objective scoring. One possible solution for the latter problem is to refer to the judgments of experts (e.g. based on a scoring manual or a rubric; cf. Gardner & Gess-Newsome, 2011) combined with an evaluation of inter-rater agreement. Both increased work intensity and time requirements for item scoring could decrease the practicability of open-ended items for large-scale studies on teachers' PCK.

*Item Stem.* An additional item development issue is associated with the amount of pedagogical context information that teachers need for an adequate response to PCK items. As teachers' PCK pertains to teaching particular topics framed by particular pedagogical contexts, item stems require pedagogical context information, for instance, students' pre-knowledge and grade, and intended learning goals for the topic. Due to space and time constraints, the pedagogical context information of PCK items is often reduced to a minimum, which makes a context-sensitive response to PCK items more difficult (Baxter & Lederman, 1999; Kromrey & Renfrow, 1991).

*Item Validity.* A further issue concerns the validity of items (Schilling & Hill, 2007), in the sense that they ensure PCK is measured rather than content knowledge or general pedagogical knowledge. One method to test the validity of PCK items is to use expert ratings (Carlson, 1990). A second option is to measure discriminant validity by using PCK items with contrasting samples of expertise on content knowledge (e.g. scientists) or general pedagogical knowledge (e.g. teachers of other subjects or pedagogues; Krauss, Baumert & Blum, 2008). Discriminant validation with nonteaching professionals may help to test that the knowledge being measured is unique for teaching professionals. A third method is to use the technique of known groups (Hattie & Cooksey, 1984). Using this technique, one criterion is that item scores or test scores discriminate between groups that are theoretically known to differ (e.g. student teachers at university, trainee teachers in practical internships, and experienced in-service teachers). Finally, a factor analysis might help to clarify the structure validity by indicating how much latent constructs are measured by a PCK test, for example, pedagogy, content, and context (cf.

Hill et al., 2008; Rohaan et al., 2009). This study focuses on declarative PCK and presents the development as well as the evaluation of a paper-and-pencil test.

## Method

### Research Design

The test item development and evaluation procedure, which was based on the classical test theory (Allen & Yen, 2002), can be divided in five steps: (I) theoretical conceptualization of test scales, (II) video analysis and item development, (III) pilot study, (IV) main study, and (V) validation study.

*Step I: Conceptualization of Test Scales.* The test construction started with a theoretical–deductive conceptualization of PCK test scales on the basis of a literature review: test scales were deduced from theoretical considerations about teachers' PCK. According to the literature review, the PCK test should cover two test scales: knowledge of student learning and conceptions (PCK I) and knowledge of representations and strategies (PCK II). The final draft of the piloted PCK test had five items in the scale PCK I and ten items in the scale PCK II. We used the PCK test to focus on biology teachers' declarative components of PCK (Baumert et al., 2004).

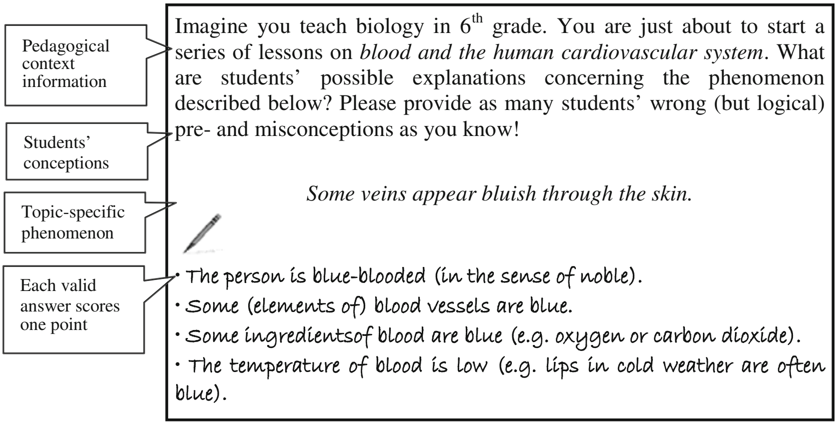
*Step II: Video Analysis and Development of Test Items.* This step involved 50 videotaped ninth grade biology lessons on the topic of blood and the human cardiovascular system, which were collected during previous studies (Jatzwauk, Rumann & Sandmann, 2008; Tiemann, Rumann, Jatzwauk & Sandmann, 2006; Wadouh, Sandmann & Neuhaus, 2009), and were reanalyzed aiming to identify frequently used models, representations, and explanations.

Beside some other research questions, the video analysis aimed to explore teachers' PCK on the topic in an empirical–inductive manner. First, the videotaped biology lessons were analyzed and broken down into different phases of instruction. Then, those situations were marked in which the teachers used topic-specific illustrations, representations, explanations, and models to teach the topic of blood and the human cardiovascular system. In addition, those situations were marked that showed students with topic-specific preconceptions and misconceptions. The most frequent illustrations, representations, models, and students' conceptions were listed. The findings of the video analysis were added to the findings of an interview study on teachers' PCK concerning the topic of the human circulatory system, as well as the findings of another interview study on students' preconceptions and misconceptions on the topic of blood and the human cardiovascular system. In that way, we generated a catalogue—of typical illustrations, representations, and models (background for PCK II), as well as students' preconceptions and misconceptions (background for PCK I) about the topic—which was derived from video-observations (Schmelzing, Wüsten, Sandmann & Neuhaus, 2008), interviews with teachers (Loughran, Berry & Mulhall, 2006), and interviews with students (Sungur, Tekkaya & Geban, 2001).

Based on this collected data from video-observations, items were developed on the two scales: related to biology teachers' PCK of student learning and conceptions (PCK I) and about their PCK of representations and strategies (PCK II). In addition, the item development was supported by experienced biology teachers to benefit from their practical teaching experiences.

*Step III: Pilot Study.* The developed PCK test (40 items) was piloted with a random sample of preservice and in-service biology teachers from North Rhine-Westphalia (Federal State of Germany) to obtain initial insights into the clarity and practicability of the developed items and to explore the psychometric characteristics of the items and test scales. All participants in the pilot study got a voucher of 15€. The participants were given 3 weeks to take the test at home. Forty-two participants (10 student teachers from the university and 32 in-service biology teachers) returned the completed test. The gathered data was used for an evaluation of internal consistency[1] of test scales, statistics for each item (item difficulties, variances, and discrimination indices), item clarity, and item practicability. The objectivity of the developed coding manual was evaluated by inter-rater agreement using unadjusted intraclass correlation coefficient ($\mathrm{ICC}_{\mathrm{unjust}}$)[2] (Shrout & Fleiss, 1979) on the basis of ten randomly selected tests. The $\mathrm{ICC}_{\mathrm{unjust}}$ indicated excellent[3] agreement of the two independent raters on the item scoring, $\mathrm{ICC}_{\mathrm{unjust}} = 0.77$, $F(9,9) = 7.50$, $p = 0.003$. Twenty-five items—which did not match the valid item difficulty[4], $0.20 < P_{\mathrm{m}} < 0.80$, also showed a low discrimination coefficient ($r_{\mathrm{it}}$),[5] $r_{\mathrm{it}} < 0.30$, or were found to be ambiguous for solving or evaluating—were rejected. The final entire PCK test consisted of 15 open-ended items in 2 scales, one scale consists of 5 items and the other scale consists of 10 items. These 15 items were used for the main study and the validation study.

Due to the advantages which allow measuring teachers' individual teaching experience, we developed open-ended items. The item stem of the open-ended items covered pedagogical context information about an imaginary learning group (with certain pre-knowledge and of a certain grade belonging to a certain type of school), a particular topic, the current status concerning the series of lessons, and intended learning goals (see Figs. 2 and 3; cf. Schmelzing, 2010). To set the participants into the role of experts, most of the PCK items were formulated as questions a preservice teacher could ask their mentor (see Fig. 3). The items asked the participant to list as many as possible students' preconceptions and misconceptions concerning one specific biological phenomenon, representation, or model (see Figs. 2 and 3). Thus, the open-ended items measured the size of a PCK repertoire instead of distinguishing between "right" and "wrong" PCK responses (Baxter & Lederman, 1999; Kromrey & Renfrow, 1991). To score the open-ended items in an objective manner, a coding manual was developed. Participants' written answers were justified as valid if they might be confirmed by comparing them to the findings of the video-observations and interview studies with students and teachers on the topic of blood and the human cardiovascular system or if they might be backed up by professional consensus (e.g. consensus found in the literature on biology education) (cf. Carlson, 1990; Rohaan et al., 2009). To incorporate teachers' individual teaching experiences, all answers which were not described in the literature, but appeared twice or more in the pilot study, were scored as valid too. In the end, valid answers were counted and summed up for an item score. In that way, every item had numerous valid answers. Hence, the PCK test was conceptualized as a power test with no score limit.

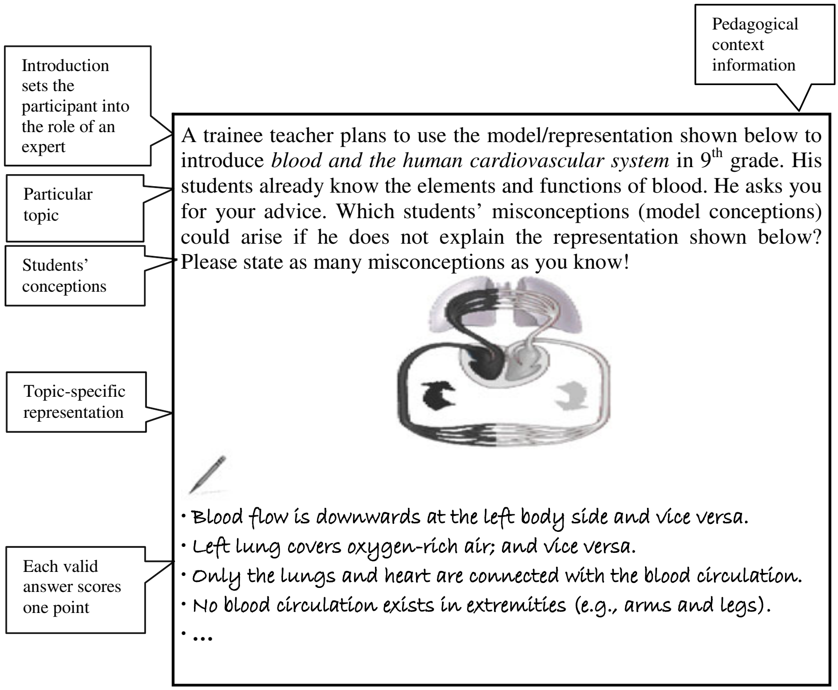

Pedagogical context information
Students' conceptions
Topic-specific phenomenon
Each valid answer scores one point

Imagine you teach biology in 6th grade. You are just about to start a series of lessons on *blood and the human cardiovascular system*. What are students' possible explanations concerning the phenomenon described below? Please provide as many students' wrong (but logical) pre- and misconceptions as you know!

*Some veins appear bluish through the skin.*

- The person is blue-blooded (in the sense of noble).
- Some (elements of) blood vessels are blue.
- Some ingredientsof blood are blue (e.g. oxygen or carbon dioxide).
- The temperature of blood is low (e.g. lips in cold weather are often blue).

*Figure 2.* Item sample for measuring biology teachers' declarative PCK (test scale PCK I) with a selection of sample answers (*below the pencil*) and explanations (*balloons on the left*) (cf. Schmelzing, Wüsten, Sandmann & Neuhaus, 2010b, in German)


Pedagogical context information
Introduction sets the participant into the role of an expert
Particular topic
Students' conceptions
Topic-specific representation
Each valid answer scores one point

A trainee teacher plans to use the model/representation shown below to introduce *blood and the human cardiovascular system* in 9th grade. His students already know the elements and functions of blood. He asks you for your advice. Which students' misconceptions (model conceptions) could arise if he does not explain the representation shown below? Please state as many misconceptions as you know!

- Blood flow is downwards at the left body side and vice versa.
- Left lung covers oxygen-rich air; and vice versa.
- Only the lungs and heart are connected with the blood circulation.
- No blood circulation exists in extremities (e.g., arms and legs).
- ...

*Figure 3.* Item sample for measuring biology teachers' declarative PCK (test scale PCK II) with a selection of sample answers (*below the pencil*) and explanations (*balloons on the left*). For the development of these items, compare Schmelzing et al., 2008, 2010a

*Step IV: Main Study.* This step involved the final PCK test that was given to a sample ($n = 93$) of student teachers (preservice teachers at the university, $n = 22$), trainee teachers (preservice teachers in teacher traineeship, $n = 22$), and experienced teachers (in-service teachers, $n = 49$). The main study was used for the evaluation of the developed PCK test, looking at item statistics (item difficulties, item variances, and item discrimination indices), internal consistency of test scales, and objectivity of item scoring. The data from the main study was also used to check the validity of the PCK test by using the technique of known groups believed to have different measureable traits (Hattie & Cooksey, 1984), by comparing the test scores of student teachers from the university, trainee teachers from practical internships, and experienced biology teachers.

*Step V: Validation Study.* Finally, the PCK test was used in a discriminant validation study with a small contrasting sample of biology graduates ($n = 12$). We evaluated whether biological content knowledge was sufficient for participants to answer the PCK items and examined whether the PCK test actually measured knowledge that was unique to the teaching profession.

## Participants

*Main Study.* An information letter was sent to 251 randomly selected schools and 35 teacher training colleges in North Rhine-Westphalia to recruit biology teachers and biology trainee teachers of all school grades. Neither the sample of the preservice teachers nor the sample of the in-service teachers was chosen from a sample of highly effective teachers. The letters covered background information about the study, promised to give individual feedback, a certificate, and 30€ for all participants in the main or the validation study. The participants were given 3 weeks to take the test at home.

Overall, 93 participants filled in the test. The sample was divided into 3 subsamples according to their teaching experience: student teachers ($n=22$; from the same university), trainee teachers ($n=22$), and in-service biology teachers ($n=49$). At the time of data collection, the 22 student teachers had, on average, participated in 5 biology education courses at the university ($SD=3$). Most of the student teachers (81 %) did not have teaching experience. Gender could not be determined because of missing data.

Twenty-two participants were trainee teachers with 68.2 % female and 31.8 % male participants. Twenty-one of them had passed the first year of a 2-year teaching internship and the remaining person had finished the second year. Most of the trainee teachers (72.7 %) taught at the upper secondary schools and 27.3 % at lower secondary schools.

Forty-nine participants were in-service biology teachers, of which 73.5 % were females and 26.5 % males. They had a mean teaching experience of 12 years ($SD=10$; within a range of 1 – 34 years). Fourteen percent of in-service biology teachers had a teaching qualification for lower secondary schools, 10 % had a teaching qualification for upper secondary schools, and 76 % had a teaching qualification for both lower and upper secondary schools.

*Validation Study.* The data of the main study ($n=93$) was also used for the validation study. The PCK test was also given to a contrasting sample of biology graduates from different universities in Germany. For this purpose, biologists were recruited with an information letter which was sent to various biology institutes. Of the 12 biologists that responded, 8 were female and 3 were male (information was missing from one participant).

### *Data Analysis*

*Main Study.* The data gathered in the main study was used to evaluate psychometric properties of the two test scales PCK I and PCK II (internal consistency and distribution of test scores), as well as to evaluate psychometric item statistics (item difficulties, item variances, and item discrimination indices). Moreover, objectivity of the developed coding manual was evaluated by inter-rater reliability. Two independent raters scored 20 randomly selected tests (overall, 300 items). Inter-rater agreement was evaluated using $\mathrm{ICC}_{\mathrm{unjust}}$ (Shrout & Fleiss, 1979).

*Validation Study.* Finally, we checked the validity of the instrument by using the technique of known groups (Hattie & Cooksey, 1984), by comparing the test scores of student teachers, trainee teachers, and experienced biology teachers. To test the significance of the mean differences between the PCK test scores of student teachers, trainee teachers, and experienced teachers, an analysis of variance (ANOVA)[6] was computed. Levene's test was used to test equality of variances in the different subsamples. Due to the unequal group sizes, Hochbergs GT2 Test was calculated as a post hoc analysis. In an additional validation study with a small contrasting sample of biologists, we compared the subsample of experienced biology teachers' mean PCK test score ($n = 49$) with the biologists' mean PCK test score ($n = 12$) using the $t$ test. The effect size, Cohen's $d$, was calculated for significant differences.

## FINDINGS

### *Main Study*

*Findings Concerning the Test Scales.* We evaluated the internal consistency of the scales of the developed PCK test by determining Cronbach's alpha ($\alpha$) values. Internal consistency was satisfactory for both the scales PCK I and PCK II, as well as for the whole PCK test, $\alpha = 0.85$ (see Table 2; cf. Schmelzing, 2010).

To evaluate the distribution of test scores or the distribution of test item difficulties, a Kolmogorov–Smirnov test was performed for each of the three subsamples (student teachers, trainee teachers, and experienced biology teachers). The Kolmogorov–Smirnov test confirmed a normal distribution of test scores/test item difficulties for all three subsamples. This result indicates that the PCK test can distinguish between high and low achievers for each of the three subsamples.

**TABLE 2**

Internal consistency of test scales with valid cases ($n$), mean scores ($M$), standard deviation ($SD$), maximum score (Max), number of items ($m$), and Cronbach's alpha ($\alpha$)

| $n$ | $M$ | $SD$ | $Max$ | $m$ | $\alpha$ |
|---|---|---|---|---|---|
| | | PCK I | | | |
| 93 | 11.74 | 4.50 | 23 | 5 | 0.75 |
| | | PCK II | | | |
| 93 | 27.05 | 8.41 | 46 | 10 | 0.80 |
| Total | 38.80 | 11.52 | 69 | 15 | 0.85 |

*Findings Concerning the Item Statistics.* The psychometric evaluation of item statistics covered the discrimination coefficient and item difficulties. All 15 items showed an acceptable discrimination coefficient ($r_{\text{it}}$), $r_{\text{it}} > 0.30$. Furthermore, all items except for one showed valid item difficulties ($P_{\text{m}}$) within an acceptable range, $0.20 < P_{\text{m}} < 0.80$.

*Findings Concerning the Objectivity.* To evaluate the objectivity of item scoring, we tested the inter-rater agreement on the basis of the scoring guide we had developed. According to Shrout & Fleiss (1979), the inter-rater agreement for the whole PCK test was excellent,[7] $\text{ICC}_{\text{unjust}} = 0.92$, $F(299,299) = 24.45$, $p < 0.001$.

### *Validation Study*

For the participants of the main study ($n = 93$), the teachers' educational level (student teachers, trainee teachers, or experienced teachers) showed a trend on the mean PCK test scores, $F(2,92) = 2.99$, $p = 0.055$. Post hoc analysis for unequal group sizes (Hochbergs GT2) indicated increasing PCK test scores from student teachers, $M = 4.38$, $SD = 1.37$, to trainee teachers, $M = 5.10$, $SD = 1.45$, and from trainee teachers to experienced biology teachers, $M = 5.34$, $SD = 1.58$. Overall, there was a significantly strong[8] difference, $p = 0.049$, $d = 0.65$, between student teachers' and experienced teachers' mean PCK test scores (see Fig. 4). Levene's test showed equality of variances in the different samples, $F(2,90) = 0.054$, $p = 0.947$.

In addition to the technique of known groups, we checked the discriminant validity of the developed PCK test, comparing the subsample of experienced biology teachers' mean PCK test score ($n = 49$) with biologists' mean PCK test score ($n = 12$). Consistent with theoretical considerations about the uniqueness of PCK for teaching

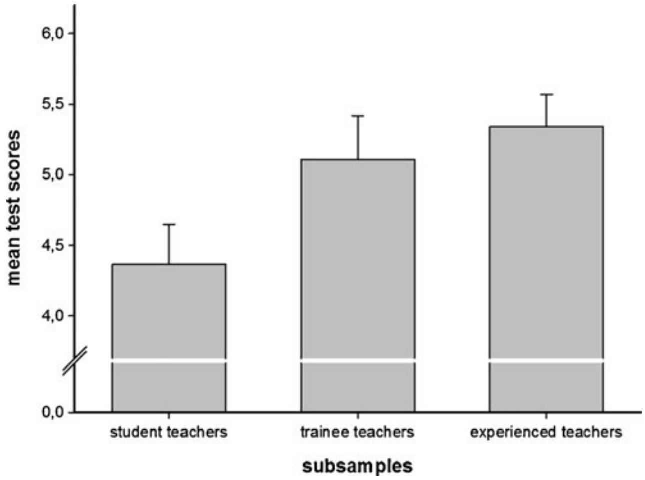


*Figure 4.* Mean PCK test scores, trend line, and error bars of student teachers ($n$ = 22), trainee teachers ($n$ = 22), and experienced teachers ($n$ = 49). There is a significant difference between student teachers and experienced teachers ($p$ = 0.049; $d$ = 0.65) (cf. Schmelzing et al., 2010, in German)

professions, a *t* test[9] showed significantly lower PCK test scores of biologists compared to experienced biology teachers (see Table 3).

## Conclusion, Critical Reflection, and Discussion

In conclusion, the reported test construction provided a first step towards a more standardized and quality-oriented measurement of PCK with a paper-and-pencil test for larger sample sizes. The test scales were objective and reliable; the item statistics were all within a valid range.

**TABLE 3**

Comparison of biologists' (b) and experienced biology teachers' (et) mean test scores (*t* tests) (cf. Schmelzing et al., 2010b, in German)

| *Subsamples* | *t test* | *Cohen's d* |
|---|---|---|
| | PCK I | |
| b, et | $t(59) = 2.00$, $p = 0.050$ | 0.53 |
| | PCK II | |
| b, et | $t(59) = 3.59$, $p = 0.001$ | 1.12 |
| Total | $t(59) = 2.83$, $p = 0.006$ | 1.00 |

The PCK test required 20 min to take, so it was time-efficient for respondents. Differences between student teachers, trainee teachers, and in-service teachers were not significant but showed an expected trend in an upward direction. Comparisons between experienced biology teachers and biologists provided empirical clues which back up the validity of the PCK test developed in the study. In any case, it indicated that content knowledge alone is not sufficient for participants to answer the developed PCK items. Therefore, the results indicate that the developed PCK test measures knowledge which is unique to biology teachers (cf. Shulman, 1986).

On critical reflection, there were several issues concerning the design and the methodology of this study that require discussion. First of all, for psychometric reasons, the number of items in each scale needs to be increased. Furthermore, limitations of the sample need to be discussed. The sample size of the main study was relatively small ($n = 93$) for an evaluation of the PCK test which aimed to produce generalizable findings. The small sample size could be a reason why the comparison between student teachers, trainee teachers, and in-service teachers showed only a trend, but no significance. In addition, the sample was limited to one federal state of Germany and the subsample of student teachers all attended a single university. Thus, generalizability about the psychometric test properties has to be interpreted with caution. The same issue also applies to the validity study with biologists which was based on a very small sample size ($n = 12$). In addition, all participants were volunteers whose motivation to participate was influenced by the fact that they received a financial reward. Thus, there is an increased probability that the participants of the main study were above average in motivation, enthusiastic, and showed a higher level of self-confidence regarding their PCK. If this was actually the case, it is arguable whether the sample was representative of all biology teachers. Finally, testing at home may also be a problem because participants may have received assistance from others.

Concerning the design and the methodology in constructing the PCK test, the open-ended item format of the developed PCK test could imply an item bias resulting in low item scores. This might be a hint that the developed PCK test could not distinguish between participants who lacked motivation and participants who lacked PCK. In addition, the test relied heavily on declarative PCK. The test could be improved by items which can measure the application of PCK to realistic classroom situations, for example, written or video simulated classroom vignettes (cf. Schmelzing et al., 2009). Not until then will the PCK test cover both the declarative and the procedural knowledge types and, therefore, the whole concept of PCK. Future studies might investigate how teachers' PCK about one topic relates to their PCK on other topics, by using multiple PCK tests on various topics and correlating the test scores (Jüttner & Neuhaus, 2012; Tepner, Borowski, Dollny, Fischer, Jüttner, Kirschner et al., 2012).

Within the context of biology teacher education, the PCK test developed in this study could be used for teacher certification as well as for teacher self-evaluation for the topic of blood and the human cardiovascular system. Furthermore, the development of PCK items by preservice and in-service teachers could be used as a method in teacher education. The development of PCK items provides an opportunity to research and structure preservice and in-service teachers' reflection on their own PCK. In future studies, it would be interesting to compare randomly chosen biology teachers to highly effective biology teachers. Such a study would provide new insights into biology teachers' PCK that is especially important in effective teaching. Furthermore, such a study would give some more hints concerning the validity of the paper-and-pencil test developed.

## Acknowledgments

This study was funded by the German Research Foundation and its full details (including figures, tables, and diagrams) were reported in a dissertation in German (Schmelzing, 2010).

## Notes

[1] Internal consistency measures whether several items that propose to measure the same general construct produce similar item scores.

[2] The ICC indicates agreement of two or more independent raters on continuous variables (in that case item scores).

[3] According to Shrout & Fleiss (1979), an ICC>0.75 indicates excellent agreement.

[4] Item difficulty describes the percentage of participants who answered an item correctly.

[5] The discrimination coefficient is a measure of the power of the item to distinguish between high and low scorers.

[6] An ANOVA is a statistical test of whether or not the means of several groups are all equal. In our case, assumptions of the ANOVA could be violated because the group sizes of student teachers, trainee teachers, and experienced teachers were unequal. To validate the findings, a nonparametric (more robust) Kruskal–Wallis test was used. The Kruskal–Wallis test showed comparable results.

[7] According to Shrout & Fleiss (1979), an ICC>0.75 indicates an excellent agreement.

[8] According to Cohen (1988), $d=0.2$ indicates a moderate effect, $d=0.5$ indicates a medium effect, and $d=0.8$ indicates a strong effect.

[9] Assumptions of the *t* test could be violated because the group sizes are unequal. For validation of findings, a nonparametric (more robust) Mann–Whitney–Wilcoxon test (*U* test) was used. The Mann–Whitney–Wilcoxon test showed comparable results.

## References

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

Allen, M. J. & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove: Waveland.

Anderson, J. R. (1981). *Cognitive skills and their acquisition*. Hillsdale: Lawrence Erlbaum.

Schmelzing, S. (2010). Das fachdidaktische Wissen von Biologielehrkräften: Konzeptionalisierung, Diagnostik, Struktur und Entwicklung im Rahmen der Biologielehrerbildung. [Pedagogical content knowledge of biology teachers: conceptualization, diagnostics, structure and development within biology teacher education]. Berlin: Logos.

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

Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. (2008). Evaluation von zentralen Inhalten der Lehrerbildung: Ansätze zur Diagnostik des fachdidaktischen Wissens von Biologielehrkräften. [Evaluation of central aspects of teacher education: Initial stages of diagnosing the pedagogical content knowledge of biology teachers.] *Lehrerbildung auf dem Prüfstand, 1*, 641–663.

Schmelzing, S., Fuchs, Ch., Wüsten, S., Sandmann, A. & Neuhaus, B. (2009a). Entwicklung und Evaluation eines Instruments zur Erfassung des fachdidaktischen Reflexionswissens von Biologielehrkräften. [Development and Evaluation of an instrument to measure the reflection abilities of biology teachers with regard to pedagogical content knowledge.] *Lehrerbildung auf dem Prüfstand, 2*, 56–80.

Schmelzing, S., Wüsten, S, Sandmann, A. & Neuhaus, B. (2010a). Measuring declarative and reflective components of biology teachers‘ pedagogical content knowledge. In: M.F. Taşar & G. Çakmakcı (Eds.). Contemporary science education research: preservice and inservice teacher education. Ankara, Turkey: Pegem Akademi, pp. 71–77.

Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. (2010b). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. [Pedagogical content knowledge and reflection in frame of biology teacher education.] *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 189–207.

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

Baumert, J., Blum, W. & Neubrand, M. (2004). Drawing the lessons from PISA 2000. *Zeitschrift für Erziehungswissenschaft, Beiheft, 3*, 143–157.

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

Baxter, J. A. & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 147–161). Dordrecht: Kluwer Academic.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A. & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: TIMMS International Study Center, Boston College.

Carlson, R. E. (1990). Assessing teachers' pedagogical content knowledge: Item development issues. *Journal of Personal Evaluation in Education, 4*, 157–163.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum.

Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. In L. Darling-Hammond (Ed.), *Review of research in education* (20th ed., pp. 3–56). Washington, DC: American Educational Research Association.

Fernandez-Balboa, J.-M. & Stiehl, J. (1995). The generic nature of pedagogical content knowledge among college professors. *Teaching and Teacher Education, 11*, 293–306.

Gardner, A. L., & Gess-Newsome, J. (2011, April). *A rubric to measure teachers' knowledge of inquiry-based instruction using three data sources*. Paper presented at the NARST Annual Meeting, Orlando.

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*, 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 Academic.

Gess-Newsome, J., Cardenas, S., Austin, B. A., Carlson, J., Gardner, A. L., Stuhlsatz, M. A. M., Wilson, C. D. et al. (2011, April). *Impact of educative materials and transformative professional development on teachers' PCK, practice, and student achievement.* Paper presented at the NARST Annual Meeting, Orlando.

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

Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of PCK. *Teachers and Teaching: Theory and Practice, 11*, 237–292.

Hattie, J. & Cooksey, R. W. (1984). Procedures for assessing the validities of tests using the "known-groups" method. *Applied Psychological Measurement, 8*, 295–305.

Heller, J. L., Daehler, K. R., Shinohara, M., & Kaskowitz, S. R. (2004, April). *Fostering pedagogical content knowledge about electric circuits through case-based professional development*. Paper presented at the NARST Annual Meeting, Vancouver. Retrieved from http://www.wested.org/understandingscience/downloads/ppr_fostped.pdf.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M. & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives, 5*, 371–406.

Hill, H. C., Loewenberg Ball, D. & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 38*, 372–400.

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

Jatzwauk, P., Rumann, S. & Sandmann, A. (2008). Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler—Ergebnisse einer Videostudie. [The influence of tasks on the students' learning performance in biology lessons—Results of a video study]. *Zeitschrift für Didaktik der Naturwissenschaften, 14*, 263–281.

Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education, 45*, 169–204.

Knight, P. (2002). A systemic approach to professional development: Learning as practice. *Teaching and Teacher Education, 18*, 229–241.

Krauss, S., Baumert, J. & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *The International Journal of Mathematics Education, 40*, 873–892.

Kromrey, J. D., & Renfrow, D. D. (1991). *Using multiple choice examination items to measure teachers' content specific pedagogical knowledge*. Annual Meeting of the Eastern Educational Research Association, Boston. Retrieved from ERIC database (ED329594).

Lee, E. & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education, 30*, 1343–1363.

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

Loughran, J., Milroy, P., Berry, A., Gunstone, R. & Mulhall, P. (2001). Documenting science teachers' pedagogical content knowledge through PaP-eRs. *Research in Science Education, 31*, 289–307.

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

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

Mulhall, P., Berry, A., & Loughran, J. (2003). Frameworks for representing science teachers' pedagogical content knowledge. *Asia-Pacific Forum on Science Learning and Teaching, 4*(2), Article 2. Retrieved from http://www.ied.edu.hk/apfslt/v4_issue2/mulhall/index.htm.

OECD (2000). *Measuring student knowledge and skills—The PISA 2000 Assessment of Reading Mathematical and Scientific Literacy*. Paris: OECD—Organisation for Economic Co-Operation and Development.

Park, S., Jang, J., Chen, Y. & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education, 41*, 254–260.

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*, 261–284.

Piburn, M., & Sawada, D. (2003). *Reformed Teaching Observation Protocol (RTOP): Reference manual* (rep. no. IN00-3). Retrieved from http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/.

Polanyi, M. (1958). *Personal knowledge*. Chicago: University of Chicago Press.

Rohaan, E. J., Taconis, R. & Jochems, W. M. G. (2009). Measuring teachers' pedagogical content knowledge in primary biology education. *Research in Science and Technological Education, 27*, 327–338.

Ryle, G. (1971). 'Knowing how and knowing that' (Proceedings of the Aristotelian Society, 1946). In G. Ryle (Ed.), *Collected papers. Volume II: Collected essays, 1929–1968* (pp. 212–225). London: Hutchinson & Co.

Schilling, S. G. & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement: Interdisciplinary Research and Perspectives, 5*, 93–106.

Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., Schwille, J. et al. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries. MT21 report.* East Lansing: Michigan State University.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Shulman, L. (1987). Knowledge and teaching: Foundations 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*, 1–20.

Staub, F. C. & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains. Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94*, 344–355.

Stillings, N., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L. & Rissland, E. L. (1995). *Cognitive science: An introduction*. Cambridge: MIT.

Sungur, S., Tekkaya, C. & Geban, O. (2001). The contribution of conceptual change texts accompanied by concept mapping students understanding of the human circulatory system. *School Science and Mathematics: Official Journal of the School Science and Mathematics Association, 101*, 1–14.

Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education: an Internal Journal of Research and Studies, 4*, 99–110.

Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Wirth, J. et al. (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.

Tiemann, R., Rumann, S., Jatzwauk, P. & Sandmann, A. (2006). Aufgaben aus Lehrersicht. [Tasks from teachers' point of view]. *Der Mathematische und Naturwissenschaftliche Unterricht, 59*, 304–307.

Wadouh, J., Sandmann, A. & Neuhaus, B. J. (2009). Vernetzung im Biologieunterricht—deskriptive Befunde einer Videostudie. [Knowledge linking levels in biology lessons—Descriptive results of a video study]. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 69–87.

Stephan Schmelzing and Stefanie Brandenbusch
*Formerly: Research Group and Graduate School "Teaching and Learning Science"*
*University of Duisburg–Essen*
*Essen, Germany*

Jan H. van Driel
*ICLON—Leiden University Graduate School of Teaching*
*Leiden, Netherlands*

Melanie Jüttner and Birgit J. Neuhaus
*Biology Education, Faculty of Biology*
*Ludwig-Maximilians-University Munich*
*Winzererstr. 45, 80797 Munich, Germany*
*e-mail: birgit.neuhaus@lrz.uni-muenchen.de*

Angela Sandmann
*Biology Education*
*University of Duisburg–Essen*
*Essen, Germany*