

Development of Items for a Pedagogical Content Knowledge-Test Based on Empirical Analysis of Pupils' Errors.

Melanie Jüttner, Birgit Neuhaus

► To cite this version:

Melanie Jüttner, Birgit Neuhaus. Development of Items for a Pedagogical Content Knowledge-Test Based on Empirical Analysis of Pupils' Errors.. International Journal of Science Education, 2011, pp.1. 10.1080/09500693.2011.606511 . hal-00724618

HAL Id: hal-00724618 https://hal.science/hal-00724618

Submitted on 22 Aug2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of Science Education



Development of Items for a Pedagogical Content Knowledge-Test Based on Empirical Analysis of Pupils' Errors.

Journal:	International Journal of Science Education		
Manuscript ID:	TSED-2011-0034-A.R2		
Manuscript Type:	Research Paper		
Keywords :	pedagogical content knowledge, teacher knowledge, secondary school		
Keywords (user):	pupils' errors, pupils' understanding, biology teachers		

SCHOLARONE[™] Manuscripts

Development of Items

for a Pedagogical Content Knowledge-Test

Based on Empirical Analysis of Pupils' Errors

Abstract

In view of the lack of instruments for measuring biology teachers' pedagogical content knowledge (PCK), this article reports on a study about the development of PCK-items for measuring teachers' knowledge of pupils' errors and ways for dealing with them. This study investigated ninth and tenth grade German pupils' (n = 461) drawings in an achievement test about the knee-jerk in biology, which were analysed by using the inductive qualitative analysis of their content. The empirical data were used for the development of the items in the PCK-test. The validation of the items was determined with think-aloud interviews of German secondary school teachers (n = 5). If the item was determined, the reliability was tested by the results of German secondary school biology teachers (n = 65) who took the PCK-test. The results indicated that these items are satisfactorily reliable (Cronbach alpha values ranged from .60 to .65). We suggest a larger sample size and American biology teachers be used in our further studies. The findings of this study about teachers' professional knowledge from the PCK-test could provide new information about the influence of teachers' knowledge on their pupils' understanding of biology and their possible errors in learning biology.

Keywords: professional knowledge; biology teachers; pedagogical content knowledge; pupils' errors; pupils' understanding

al c

Introduction

Decades ago, Shulman (1986; 1987) originally defined the professional knowledge of teachers by using several categories. Since then, many research groups have tried to operationalise teachers' professional knowledge (Abell, 2007; Baumert et al., 2010; Park & Oliver, 2008) for the development of measurement instruments. Even now, teachers' professional knowledge, especially the pedagogical content knowledge (PCK), is not easy to define and measure. This article presents a study to develop an instrument for measuring PCK on the basis of empirically analysed pupils' errors. The three steps of the development in this project will be described: (1) analysis and categorisation of the pupils' answers to an achievement test; (2) development of PCK-items according to knowledge about pupils' errors based on the results of step one and determination of the validity of the items; and (3) analysis of the reliability and objectivity of the P. P. developed PCK-test.

Theoretical Background

Professional Knowledge of Science Teachers

In 1987, Shulman proposed seven categories of teachers' professional knowledge. Numerous researchers in the field of science education still use this work to organise research efforts or to design and carry out research studies (e.g., Baumert et al., 2010; Hashweh, 2005; Hill, Ball, & Schilling, 2008; Lee & Luft, 2008; Loughran, Mulhall, & Berry, 2008; Park, Jang, & Chen, 2010; Schmidt et al., 2007; Tatto et al., 2007; van Driel, Beijaard, & Verloop, 2001). For example, Baumert et al. (2010) helped organise their evaluation of teachers' professional knowledge by using three categories: pedagogical content knowledge (PCK), content knowledge (CK) and pedagogical knowledge (PK). Over the past decade, many research projects have used these three main categories of teachers' professional knowledge to provide a theoretical background. Unlike most of the current studies on mathematics teachers (e.g., Abell, 2007; Baumert et al., 2010; Hill & Ball, 2004; Lipowsky, 2006; Park & Oliver, 2008; Schmidt et al., 2007), the project *ProwiN* (the German acronym for *Professional Knowledge in Science*) (Fischer, Borowski, & Tepner, in press) is a first step to close the gap in research about teachers' professional knowledge in science education (Abell, 2007). Three cooperating German Universities analysed the PCK, the CK and the PK of science teachers based on a theoretical model (Tepner et al., submitted) and the results of researchers-developed test instruments. The study presented here is part of the project *ProwiN* with a focus on one category of the professional knowledge—the PCK of biology teachers.

Theoretical Conceptualisations of the Pedagogical Content Knowledge (PCK)

Most studies on PCK focus on one of the following PCK components: "knowledge about pupils' understanding", "curricular knowledge", "knowledge about instructional strategies and representation" (cf. Kind, 2009; Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008; van Driel, Verloop, & de Vos, 1998). Illustrated here is the development of an instrument for measuring biology teachers' PCK with a special focus on one PCK component—the pupils' errors and their understanding of biology. Of special interest to this study is the *MT21*-study (Schmidt et al., 2007) which analysed two sub-competencies of mathematic teachers' diagnostic competency: recognition of pupils' misconceptions and criteria-guided assessment of pupils' solutions to problems (Schwarz, Wissmach, & Kaiser, 2008).

International Journal of Science Education

The PCK in this study (part of the project *ProwiN*) was generally defined according to Shulman (1986) as knowledge of structuring and describing the teaching of the content that would be necessary for preparing the subject matter so that it is comprehensible for pupils (Fischer, Borowski, & Tepner, in press). The most often used PCK components in the literature—knowledge about pupils' understanding, instructional strategies and representations (Park & Oliver, 2008; Schmelzing et al., 2010)—were chosen for the PCK-model of the project (Tepner et al., sub.).

The knowledge about pupils' understanding will be highlighted. This PCK component is defined throughout pupils' errors as well as the knowledge about the thinking of the pupils in solving problems. In addition to the three PCK components of the project (pupils' errors, models and experiments), three knowledge dimensions were defined: declarative, procedural and conditional knowledge (Alexander, Schallert, & Hare, 1991; de Jong & Ferguson-Hessler, 1996; Paris, Lipson, & Wixson, 1983). These knowledge dimensions form the required cognitive aspect of teachers' professional knowledge—here the PCK conceptualisation (Hill, Ball, & Schilling, 2008). Moreover three different biology topics that teachers have to know about—neurobiology, plants and vertebrates—were selected from the curriculum of German schools for the development of the PCK-items. In this article, we use the human reflex arc to demonstrate the idea of developing PCK-items concerning "pupils' errors" and the three different knowledge dimensions (declarative, procedural and conditional).

Pupils' Errors

In science education, there are many definitions of different terms which have been used for describing pupils' knowledge (Taber, 2009). Sometimes, some terms could be used synonymously and sometimes one and the same term is discussed in different ways. Table 1 shows an overview of the exemplary literature on some of the accepted terms about pupils' knowledge in science that were categorised and combined into two different groups for defining "pupils' errors" in our study.

[Insert Table_1 about here]

Conceptions and prior knowledge. In the literature about pupils' scientifically incorrect conceptions, two main aspects have become apparent—pupils can bring their ideas with them to the classroom from their everyday life or they can develop their ideas during the lesson and/or because of the topic being taught in the lesson. Accordingly, the terms can be classified into two groups. (1) *Conceptions*—defined as a way of pupils' thinking which they established on their own depending on their experience in everyday life but not consistent with teachers' and scientists' ideas—have often been called *misconceptions*, *preconceptions*, *alternative conceptions*, *alternative frameworks* or *children's science* (Treagust, 1988). (2) *Prior knowledge* (Allen, 2005)—described as knowledge before pupils learn something about a topic—has been studied since the 1970s. *Knowledge* is used as a tightly-defined term and *idea*, *belief*, as well as *concept* are sometimes used as synonyms (Taber, 2009). But knowledge differs from conceptions, because knowledge of pupils can be described as a set of conceptions.

Pupils' errors. Concurrent to studies on "conceptions or misconceptions", research in pedagogical psychology explores "*pupils' errors*" (Oser, Hascher, & Spychiger, 1999; Seifried & Wuttke, 2010; Spychiger, 2008). Pupils' errors are the out coming signals of their difficulties or

misunderstandings in lessons. The "errors" are different from the "misconceptions", because misconceptions could be defined as scientifically incorrect knowledge about something which can and should influence pupils' conceptual learning in the lesson (Tanner & Allen, 2005). The meaning between neutral and non-neutral terms can be distinguished, for example, "error" or "knowledge" (Taber, 2009), which indicates the possibility of saying if something is wrong or right according to the current scientific research on the topic. But there are problems within the definition as well—what is wrong and what is right in science (Cokelez, Dumon, & Taber, 2008)? Using these non-neutral—often negatively understood—terms, researchers have to define them clearly for the terms not being misunderstood (Oser, Hascher, & Spychiger, 1999). On the contrary, a neutral term will be, for example, "children's ideas" and "alternative conceptions" (Taber, 2009), which are defined more openly to include most ideas of pupils.

In this study, the term "errors" was used because we were interested in errors of the ninth and tenth grade pupils after their learning process in biology in the ninth grade. The focus was on errors of pupils because these errors could be seen as an indicator of the effectiveness of instruction (Seidel & Shavelson, 2007). Pupils' errors are an important component of teachers' PCK on how to handle errors and to examine what might be the reasons for such errors (Baumert et al., 2010; Hill, Rowan, & Loewenberg Ball, 2005; Lipowsky, 2006; Seifried, & Wuttke, 2010).

Aims

Based on the theoretical background, the following were the research questions and aims of our study:

(1) What are the typical pupils' errors about the reflex arc of the knee-jerk?

- (2) How can items for a PCK-test be developed on the basis of the identified pupils' errors to measure teachers' PCK of pupils' errors and ways for dealing with them?
- (3) Are these PCK-items reliable and valid for assessing the PCK of pupils' errors?

Method

The three parts of the study described here are: pupils' errors analysis, development of the PCKitems as well as the qualitative validation and quantitative analysis of the developed PCK-items.

Participants

An achievement test was given to 461 ninth- and tenth-grade pupils in secondary schools in Germany—204 male pupils (42.7%) and 249 female pupils (52.1%) (25 pupils without gender information) with an average age of M = 15.14 years (33 pupils without age information). All the pupils who took the achievement test (about the reflex arc) had learned about the topic before the test in their ninth grade.

The results of this achievement test were used to develop PCK-items about pupils' errors. To determine the content validity of these items, five German biology teachers—three teachers are teaching at the higher secondary school and two at the general-education secondary school— were interviewed. They had an average age of 43.8 years and an average teaching experience in biology of 13.40 years. Additionally, all the developed items concerning three different biology topics and the PCK about pupils' errors were piloted with a larger sample of 65 biology teachers. They had an average age of 39.54 years and an average teaching experience of 10.50 years (not including the two years of practicum during the apprenticeship to become a teacher). All the

participating teachers—58.5% female and 41.5% male biology teachers from six different federal states of Germany—answered the PCK-test. Of these participating biology teachers, 72.3% taught at the higher secondary schools and 27.7% taught at the general-education secondary schools.

Achievement Test

In one item of the pupils' achievement test, the pupils had to mark on a drawing of a given human body, the pathway of the impulse that travels during a *knee-jerk*—the idea was based on Hammann (2003). The pupils had about five minutes for this item (Figure 1). The expected answer of the pupils would be the pathway of an *impulse* starting at the *quadriceps muscle*, travelling to and through the *white* and *gray matter* of the *spinal cord*, and ending in the muscle again. The pathway from the muscle to the gray matter is the *sensory neuron* and the pathway back to the muscle is the *motor neuron*, which should be labelled as *afferent* and *efferent nerves*.

[Insert fig_1 about here]

Development of the PCK-items

Based on the empirically identified and counted error categories of the pupils' answers to the presented item (see Figure 1), PCK-items for biology teachers were developed in three different knowledge dimensions: declarative knowledge, procedural knowledge and conditional knowledge.

Figure 2 shows an example of three items—about the biological topic "reflex arc"—as part of the PCK-test for analysing pupils' errors with respect to the three different knowledge dimensions:

Item a) on declarative knowledge; Item b) on conditional knowledge; and Item c) on procedural knowledge. The most frequently counted pupils' errors category (i.e., category 1 in the results of this study, see Table 2) was used for the development of this item.

[Insert fig_2 about here]

At the beginning of the three items (a, b, c), the teachers got pedagogical background information (Baxter & Ledermann, 1999) about the pupils: their grade, the point in time of their achievement test and about how the test looked like. To develop PCK-items which are almost independent of the CK, the pupils' errors were written explicitly next to the drawing. Because some of the pupils' errors can be caused by materials given to them in the lesson, this is one aspect asked by Item 1b). Major aspects of teachers' PCK include not only the knowledge about possible reasons for pupils' errors but also the knowledge about pupils' upcoming errors and the use of these for planning future lessons or knowing how someone might react in special error situations (Seifried & Wuttke, 2010). These aspects are illustrated by the first (Item a) and the third item (Item c), in Figure 2. The idea of the item development is theoretically based on the empirical result that the combination of high diagnostic competency, high teachers' judgement accuracy and high frequency of instructional strategies forms the key for improving instructional quality and therefore pupils' learning outcomes (Helmke & Schrader, 1987).

The coding manual to rate the PCK-items was written on the basis of the literature in science education, the empirically identified error categories and additional results of schoolbook analysis. The answers of the teachers to Item a) were coded according to the empirical results of

the first part of the study by the use of partial credits for exactly correct answers (1), correctly given frequencies in the interval of $\pm 10\%$ of the correct answer (0.5) and other answers (0). For coding of answers to Item b), the number of meaningful named reasons was counted. Results of the document analysis (schoolbooks; school material) as well as literature (Schumacher, 2008; Seifried & Wuttke, 2010) formed the basis of this coding manual. The idea behind this item was that the more meaningful reasons the teachers name, the more the teachers might be able to consider pupils' errors in lesson planning. This coding manual was developed concerning the qualitative document analysis (schoolbooks; lesson materials; everyday life material like advertisements). Of course, if a teacher did not name a meaningful reason in the list of examples of reasons in the coding manual, this was counted with one point as well.

The possible reactions of teachers to pupils' errors in the described situation in the biology lessons (Item c) were coded with one point per each named reaction. Therefore, the listed possible reactions of the teachers are summarised in four main categories based on theoretical background: (1) error is used as orientation for further lessons; (2) error is used for other pupils so that they could learn from others errors' (analysis of errors); (3) error corrected by the pupil her-/himself so that she/he can learn from her/his errors; and (4) teachers being flexible and describe the misunderstood topic another way around during the lesson (Jüttner & Neuhaus, 2010; Caspary, 2008; Oser & Spychiger, 2005; Taber, 2009).

All the items concerning "neurobiology" and all PCK components (pupils' errors, models, experiments) were validated in an additional study. Similar items concerning the other biology topics (plants and vertebrates) were piloted. For the component "pupils' errors" there are always three PCK-items per topic according to the knowledge dimensions. Additionally, for the PCK component "model", always one item per topic—about the reasons of pupils' errors related to the

model and its use—could be added to the PCK-items about pupils' errors. So all in all, 12 PCKitems concerning the knowledge about pupils' errors were created.

Data Analysis Methods

The categories of the pupils' errors were identified using the inductive qualitative analysis of content for 10% of the pupils' answers (Hsieh & Shannon, 2005; Mayring, 2000). The inductively generated categories were described in the coding manual with which multiple counting was possible. After the coding of pupils' answers to the shown item by two independent raters, the relative frequency for each category was computed by using the statistical software SPSS 19.0.0 (2011). To analyse the interrater reliability for 10% of the sample (n = 46) the AC1 statistic (Gwet, 2002; 2008) was calculated in addition to the commonly used Cohen's kappa value (Landis & Koch, 1977; Shoukri, 2004). This was caused by the paradoxical kappa values because of the seldom coded categories (Gwet, 2002). Additionally, ten typical German schoolbooks were analysed to identify information about possible reasons for the categorised pupils' errors.

The content validity of the developed PCK-items had been addressed throughout the process of the development—the empirical data about pupils' errors were used for the creation of the items of the same biological topic. On the other hand, congruent think-aloud interviews were used which implicates that the teachers had to communicate aloud their thoughts and simultaneously write them down on the paper-and-pencil test (Ericsson & Simon, 1993; Leighton, 2009). After the verbatim transcription of the interviews, the two independent raters coded the interview transcripts by using a developed coding manual based on the theoretical model (Tepner et al.,

sub.) underlying the items. Therefore, the intraclass correlation coefficient (ICC) of the two independent raters was calculated (Field, 2009; Shrout & Fleiss, 1970).

The reliability of the PCK-test was analysed by using the Cronbach's alpha coefficient (Field, 2009). The objectivity was tested by using the intraclass correlation coefficient (ICC) concerning the coding of teachers' answers to the open-ended PCK-test items (Field, 2009).

Results

Categories of Pupils' Errors (based on pupils' drawings about the reflex arc)

Seven main categories of pupils' errors were identified and counted in the pupils' (n = 461)answers. Table 2 shows an overview of the frequencies of the seven categories. Some of the most often identified error categories are discussed with the following three examples of pupils' answers (see Figures 3, 4 and 5).

[Insert Table_2 about here]

[Insert fig_3 about here]

Figure 3 shows an example of a pupils' answer for four different categories of errors. The most frequent category 1-the drawn impulse reaches the brain (63%)-was identified in this example. Additionally, two aspects of the second category were identified: beginning and/or ending of the nerve fibres are not located at the muscle, but at bones (42%); and beginning and/or ending of the nerve fibres are not located at the muscle, but in the heel or at the tibia (14%). Also this example could be used to illustrate the fourth category: *the clearing up centre of the nerves is located in the brain* (26%). The final category identified in this example was category 5: *nerves are located in the body (stomach) and not in the spinal cord* (17%). In the drawing, a dot was used by the pupil to represent the change-over point of the pathway in the brain.

[Insert fig_4 about here]

In contrast to the example in Figure 3, the most frequently identified category 1 was not found in the pupil's answer in Figure 4. In the drawing, the pupil did not connect the brain to the reflex arc of the knee-jerk. Here, category 2b was identified: *beginning and/or ending of the nerve fibres are not located at the muscle, but at bones* (42%). Moreover, there was *no direct connection between the two nerve fibres so that the reflex arc is not closed*, that is an example for category 6 (15%).

[Insert fig_5 about here]

To show examples for every category of pupils' errors, here is the final sample of another pupil's answer (see Figure 5), where the remaining categories of errors were identified. First of all, the most frequent category 1 was identified. Furthermore, *the starting- and ending-point of the nerve fibres is in the front of the patella* (category 2a; 47%) as well as category 3: *only one way for the impulse illustrated by one single nerve* (34%) were identified. At last, the least frequent category 7 could be identified: *nerves are not named according to their function, but to their absolute location (brain = afferent nerve; muscle = efferent nerve)* (7%).

For the interrater reliability of this coding, the Cohen's kappa (κ) for all the categories was .67 ($p \le .001$, n = 462 categorisations). Because of the distribution of the data, the kappa coefficients for five categories of errors were very low (e.g., for category 2a, $\kappa = .35$), negative or even not computable; therefore, the AC1 statistic is described additionally for each category in Table 3 (Gwet, 2008; Schori, Kersten, & Abderhalden, 2006).

[Insert Table_3 about here]

The AC1 statistic (Gwet, 2008) showed values from moderate (AC1 (2a) = .50) to almost perfect (e.g., AC1 (3) = 1) interrater agreement between the two independent raters concerning all the seven pupils' error categories (see Table 3).

Validation of the Developed PCK-items

The results of the analysis of the coding of the teachers' think-aloud interview transcripts could demonstrate how often the teachers thought about or used a knowledge dimension to answer the given PCK-items (see Table 4).

[Insert Table_4 about here]

For example, to answer the third item dealing with conditional knowledge on pupils' errors (PEcond), the five teachers used, on the average, in 90.1% of their thoughts the conditional knowledge on pupils' errors. The intention of this item was to get information about teachers' conditional knowledge about pupils' errors. Additionally, while answering this item, they needed

or used, on the average, 33% of the declarative knowledge about pupils' errors as well as 16.7% of declarative content knowledge.

The interrater reliability of the two independent raters for the coding of the think-aloud interview transcripts was found to be significantly high (*ICC* _{unjust} = .99; *F* _{439,439} = 143.15; *p* < .001).

More Quality Criteria of the PCK-items.

For the internal consistency of the PCK-test scale "declarative knowledge" and the PCK component "pupils' errors", a very low Cronbach's alpha value ($\alpha = .06$; 3 items; n = 65) was found. For the two other knowledge dimensions, the Cronbach's alpha values was $\alpha = .60$ and $\alpha = .67$ (n = 65) (see Table 5).

[Insert Table_5 about here]

Finally, the intraclass correlation coefficient (ICC) of the two independent raters for the openended PCK-items was found to be statistically significant (*ICC* _(unjust) = .80; $F_{588,588}$ = 5.10; p < .001).

Discussion and Conclusions

The seven categories of pupils' errors—identified and analysed in their drawings in the achievement test item about the pathway of the impulse of the reflex arc—could be used to analyse pupils' errors, but only for this specific topic. In mathematics, ideas of general error categories exist (e.g., Ashlock, 2005), but not yet in biology. The seven error categories of ninth

International Journal of Science Education

and tenth grade German pupils reported in this article might be summed up as three main categories which might be transferred to other biological topics as well: *wrong locations of the structure; errors concerning the function* and *wrongly used terms and definitions*. The results of this study showed that most of the analysed pupils' error categories concerning the knee-jerk were usable for the categorisation of pupils' answers in this topic and that the interrater agreement (AC1) of the categories ranged from moderate to almost perfect (Gwet, 2002; Landis & Koch, 1977).

All in all, this study was able to show one way of developing PCK-items based on categorisation of pupils' errors which might enable researchers to develop their topic-specific, practical-orientated and empirically based PCK-items based on this idea. The quality of the developed PCK-items was validated by qualitative think-aloud interviews and the quantitative pilot phase. The results of these analyses demonstrated the development of a valid, reliable and objective PCK-test instrument for testing teachers' PCK concerning pupils' errors in biology.

The developed items in the PCK-test have curricular content validity because they were developed after curricular reviews and many meetings with experienced teachers. Additionally, the empirical data of the pupils' errors gave more content validity. Finally, the think-aloud interviews demonstrated that teachers mostly thought of or used the kind of knowledge (declarative, procedural or conditional) for answering the item which was intended to be measured by the respective item. The ICC of the interview transcript coding showed a satisfactory interrater agreement between the two independent raters (Field, 2009). In the end of all the interviews, four follow-up questions about the items were also asked. One of these questions asked the teacher if their answers to the items might represent their personal way of enactment in classroom situations or their way of thinking concerning example models in biology

lessons. This could be seen as face validity, because all the interviewed teachers said something similar to what Teacher K said: "My answers represent what I strive to do".

The internal consistency of the PCK-test (its Cronbach's alpha values) showed satisfactory results except for the declarative knowledge of pupils' errors—about teachers' prediction of how often the given pupils' error might arise after the learning process. One reason might be that teachers often wrongly predict their pupils' learning outcomes and because they have over-expectations of their pupils' learning outcomes (Helmke & Schrader, 1987; Hosenfeld, Helmke & Schrader, 2002).

The objectivity of the PCK-items was also confirmed with an acceptable intraclass correlation coefficient (ICC).

Due to the small sample size of this study for more detailed statistical analysis (e.g., factor analysis), further studies with samples of about 200 teachers will be conducted to test if all the developed items conform to the three dimensional theoretical model of the current project *ProwiN* (Tepner et al., sub.). Additionally, the items were translated into English and will be given to American teachers in our further studies. The results of the future think-aloud interviews might show the possible use and transferability of this study's development ideas to other countries and cultures.

Teachers' knowledge about pupils' possible errors—which may arise in a lesson or about possible reasons for these errors—is important for teachers because this knowledge has different advantages for teachers. On the one hand, it can be used as an orientation during the planning process of lessons (Schumacher, 2008; Seifried &Wuttke, 2010); and meaningful arrangement of materials, figures and models can also result. On the other hand, the information can be used for

International Journal of Science Education

the training of pupils' error analysis; pupils have the possibility to learn the so-called *negative knowledge* (Oser, Hascher, & Spychiger, 1999). In the analysis of errors made by anonymous pupils, the pupils have to change their point of view to that of a corrector. This change is the best opportunity for learning from errors (Schumacher, 2008). Therefore, empirically analysed and categorised pupils' errors, as reported in this article, are important for teachers. In addition, these results can be used for test items construction in science education research as well as in teacher education.

To conclude, the method of developing the PCK-items demonstrated in our study was able to test teachers' knowledge by its connectedness to praxis, as well as to relate pupils' performance in achievement tests directly to teachers' knowledge about their pupils' errors. So it will be possible to construct tests for pupils and teachers for analysing the correlation between pupils' knowledge and teachers' PCK concerning pupils' understanding.

As for research, the aim will be to analyse the PCK of teachers and to see if their knowledge correlates with how they prepare and teach the subject matter in different biological topics as well as with their job experience. So far, mathematical pre-service teachers' experience has been analysed in longitudinal studies (e.g. Tatto et al., 2008). Additionally, there exist a few qualitative studies with a smaller sample size evaluating the development of teachers' professional knowledge (Henze & van Driel, 2009; Friedrichsen, Abell, Pareja, Brown, Lankford, &Volkmann, 2007). In research there are different results concerning the influence of teachers' pedagogical knowledge was supported by experience but the PCK development was not. Further research concerning the analysis and conceptualisation of general pedagogical knowledge is

needed (Friedrichsen et al., 2007; Blömeke, Felbrich, & Müller, 2008). As well, biology teachers' professional knowledge development (longitudinal) and the relationship between the developments of the different components (PCK, CK, PK) could be focused on in the future to close the gap for biology teachers' research (Cohen & Yarden, 2009; Abell, 2007).

References

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105-1149). New Jersey: Lawrence Erlbaum Associates.
- Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61 (3), 315-343.
- Allen, M. (2005). Encouraging pupil error may promote better understanding of a scientific concept. *School Science Review*, 86, 79–85.
- Allen, M. (2010). Learner error, affectual stimulation, and conceptual change. *Journal of Research in Science Teaching*, 47 (2), 151-173.
- Ashlock, R. B. (2005). Error patterns in computation: Using error patterns to improve instruction (9th Edition). New-York: Macmillan.
- Bainbridge, R. (1981). To err is human: Towards a more positive approach to young children's mistakes in arithmetic. *Mathematics in School*, 10(5), 10–13.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers' mathematical knowledge, cognitive cctivation 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.

Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des

erziehungswissenschaftlichenWissens angehender Lehrkräfte [Assessment of student teachers' pedagogical knowledge]. In S. Blömeke, G. Kaiser & R. Lehmann (Eds.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 171–194). Münster, Germany: Waxmann.

Bouvier, A.(1987). The right to make mistakes. For the Learning of Mathematics, 7(3), 17–25.

- Caleon, I. S., & Subramaniam, R. (2010). Do pupils know what they know and what they don't know? Using a Four-Tier diagnostic test to assess the nature of pupils' alternative conceptions. *Research Science Education*, 40, 313-337.
- Chi, M. T. H., Slotta, J. D., & DeLeeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.
- Cokelez, A., Dumon, A., & Taber, K. S. (2008). Upper secondary French pupils, chemical transformations and the "Register of models". *International Journal of Science Education*, 30 (6), 807-836.
- Cohen, R. & Yarden, A. (2009). Experienced Junior-High-School teachers' PCK in light of a curriculum change: "The cell is to be studied longitudinal". *Research of Science Education*, 39, 131-155.

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105 - 113.

- de Laughter, J. E., & Stein, S. (1998). Preconceptions about earth science among pupils in an introductory course. *Eos*, 79, 429-434.
 - Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629-646). New York, London: Routledge.
- Driver, R.P. (1981). Pupils' alternative frameworks in science. *European Journal of Science Education*, 3(1), 93–101.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: The MIT Press. Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.

- Fischer, H.E., Borowski, A., & Tepner, O. (in press). Professional knowledge of science teachers;In B. Fraser; K. Tobin, & C. McRobbie (Eds.). Second International Handbook of Science Education, New York: Springer.
- Friedrichsen, P., Abell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46, 357-383.
- Gilbert, J., Osborne, R., & Fensham, P. (1982). Children's science and its consequences for teaching. *Science Education*, 66, 623-33.
- Gwet, K. L. (2008). *Interrater Reliability*. Wiley Encyclopedia of Clinical Trials, Copyright 2008 John Wiley & Sons, Inc.
- Gwet, K. L. (2002). Kappa Statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods For Inter-Rater Reliability Assessment*, 1, 1-5. Retrieved from http://agreestat.com/research_papers/kappa_statistic_is_not_satisfactory.pdf [Date accessed: 25 May. 2011].

Hammann, M. (2003). Aus Fehlern lernen. [Learning by errors], Unterricht Biologie, 287, 31-35.

- Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of PCK. *Teachers and Teaching: Theory and Practice*, 11, 273–292.
- Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3, 91-98
- Helm, H. (1980). Misconceptions in physics amongst South African pupils. *Physics Education*, 15, 92-7.
- Henze, I., Driel van, J. H., & Verloop, N. (2008). Development of experienced science teachers' pedagogical content knowledge of models of the solar system and the universe. *International Journal of Science Education*, 30(10), 1321-1342.
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35 (5), 330-351.
- Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge:Conceptualizing and measuring teachers' topic-specific knowledge of pupils. *Journal for Reserach in Mathematics Education*, 39, 372-400.

Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichtsund lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Lehr-Lern-Studie SALVE. [Diagnotic competency: teaching and learning relevant to student characteristics and the teachers' prediction in the teaching-learning study SALVE], In M. Prenzel & J. Doll (Eds.),*Bildungsqualität von Schule: Schulische und außerschulische* Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. Zeitschrift für Pädagogik, 45. Beiheft (pp. 65-82). Beltz: Weinheim.

- Houssart, J., & Weller, B. (1999). Identifying and dealing with misconceptions and errors in primary mathematics: Student teachers record their experiences. *Mathematics Education Review*, 11, 46-58.
- Hsieh, H.-F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Jüttner, M., & Neuhaus, B. (2011). Vom Schülerfehler zum PCK Test. Ansätze zur Entwicklung von PCK – Items. [A PCK-test based on students' errors. Rudiments for developing PCK items.] In U. Harms & I. Mackensen-Friedrichs (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik. Band 4* (pp. 27-39).Studien Verlag: Innsbruck.
- Jüttner, M. & Neuhaus, B. (2010). Using empirically analyzed pupils' errors to develop a PCK-Test. In M.F. Taşar & G. Çakmakcı (Eds.), *Contemporary science education research: preservice and inservice teacher education* (pp. 331-340). Ankara, Turkey: Pegem Akademi.
- Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education*, 45(29), 169-204.
- Krüger, D. (2007). Die Conceptuale Change-Theorie. [The conceptual change theory]. In D.Krüger & H. Vogt (Eds.), *Handbuch der Theorien in der biologiedidaktischen Forschung* (pp. 81-92). Berlin: Springer Verlag.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159–174.
- Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, 30, 1343-1363.

- Leighton, J. P.(2009). Two Types of Think Aloud Interviews for Educational Measurement:Protocol and Verbal Analysis. *Paper presented at the National Council on Measurement in Education (NCME)*, CA.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. [It depends on the teacher. Empirical evidences of the connection between teachers' competence, teachers' acting and the learning of the pupils], *Zeitschrift für Pädagogik*, 52 (51), 47-70.
- Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education*, 30 (10), 1301-1320.
- Magnusson, S., & Krajcik, J. & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In: Gess-Newsome & Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-131). Dordrecht. Kluwer Academic Publishers.
- Mayring, P. (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, North America, 1. Available at: http://www.qualitativeresearch.net/index.php/fqs/article/view/1089/2385. [Date accessed: 26 Oct. 2010].
- Nebah, S. J. (2002). Do people learn from mistakes? A pedagogical study and analysis of autobiographies and interviews. Dissertation am Departement Erziehungswissenschaften der Universität Fribourg.
- Novak, G.S. (1977). Representations of knowledge in a program for solving physics problems. Proceedings of the Fifth International Joint Conference on Artificial Intelligence. Cambridge, MA: MIT Press.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie der "negativen" Wissens. [Learning out of errors. About the psychology of the "negative" knowledge]. In W. Althof (Ed.), *Fehlerwelten: vom Fehlermachen und Lernen aus Fehlern* (pp. 11-43). Opladen: Leske + Budrich.

- Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur*.[Learning is painful. About the theory of the negative knowledge and about the practice of the culture of errors]. Weinheim: Beltz.
- Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. Contemporary *Educational Psychology*, 8, 293-316.
- Park, S., Jang, J.-Y., & Chen, Y.-C. (2010). Is Pedagogical Content Knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, 41, 245–260. doi: 10.1007/s11165-009-9163-8
- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a conceptual tool to unterstand teachers as professionals. *Research in Science Education*, 38, 261-284.
- Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Measuring declarative and reflective components of biology teachers' pedagogical content knowledge. In M. F. Tasar & G. Cakmakci (Eds.), *Contemporary science education research: Teaching* (pp. 71-77).
 Ankara, Turkey: Pegem Akademi.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L. et al. (2007). *The Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries. Mathematics Teaching in the 21st Century (MT21)* (Rep. No. MSU Center for Research in Mathematics and Science Education).
- Schori, E., Kersten, B. & Abderhalden, C. (2006). Wie reliabel ist der Fragebogen zur Analyse der Pflegedokumentation IzEP-Dok© im Instrument zur Erfassung von Pflegesystemen IzEP©? [How Reliable is the Questionnaire for the Analysis of Nursing Documentation

IzEP-Dok© as part of the Measurement Instrument for Collecting Data of the Care Systems IzEP©?], *Printernet Angewandte Pflegeforschung*, 8, 1-9

- Schumacher, R. (2008). Der produktive Umgang mit Fehlern. Fehler als Lerngelegenheit und Orientierungshilfe. [The productive handling with errors. Errors as learning opportunity and for orientation]. In R.Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp. 7-11). Freiburg: Herder.
- Schwarz, B., Wissmach, B., & Kaiser, G. (2008). "Last curves not quite correct": diagnostic competences of future teachers with regard to modelling and graphical representations. *ZDM The International Journal on Mathematics Education*, 40(5), 777-790.
- Seifried, J., & Wuttke, E. (2010). Student errors: how teachers diagnose them and how they respond to them. *Empirical Research in Vocational Education and Training (ERVET)*, 2(2), 147-162.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Education Research*, 77, 454-499.

Shoukri, M. M. (2004). Measures of interobserver agreement. Boca Raton: CRC PR Inc.

- Shulman, L. (1986). Those who understand. Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the New Reform. *Havard Educational Review*, 57, 1-22.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86 (2), 420–428.
- Spychiger, M. (2008). Ein offenes Spiel. Lernen aus Fehlern und Entwicklung von Fehlerkutur. [An open play. Learning by errors and the development of the culture of errors]. In R.

Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp. 25-48). Freiburg: Herder.

- Taber, K. S. (2009). *Progressing science education: Constructing the scientific research programme into the contingent nature of learning science*. Dordrecht: Springer.
- Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: from a scholarly approach to teaching to the scholarship of teaching. *Cell Biology Education*, 4, 1-6.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

- Tepner, O., Borowski, A., Fischer, H.E., Jüttner, M., Kirschner, S., Leutner, D. et al. (submitted). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. [Theoretical model for the developement of testitems to collect data about professional knowledge of science teachers], Zeitschrift der Naturwissenschaften.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10 (2), 159-169.
- Tsai, C.-C., & Chou, C. (2002). Diagnosing pupils' alternative conceptions in science. *Journal of Computer Assisted Learning*, 18, 157–165.

Tsamir, P., Rasslan, S., & Dreyfus, T. (2006). Prospective teachers' reactions to Right-or-Wrong tasks: The case of derivatives of absolute value functions. *Journal of Mathematical Behavior*, 25, 240–251.

van Driel, J. H., Verloop, N., & Vos, W. d. (1998). Developing science teachers' Pedagogical Content Knowledge. *Journal of Research in Science Teaching*, 35, 673-695.

van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. Journal of Research in Science

 Ministry of Education and Research; it is a cot,

 a in education' (0UH0904).

 Teaching, 38, 137-158.

* the presented project is funded by the Federal Ministry of Education and Research; it is a cooperation project embedded in the framework program of 'empirical research in education' (01JH0904).

Development of Items

for a Pedagogical Content Knowledge-Test Based on Empirical Analysis of Pupils' Errors

Abstract

In view of the lack of instruments for measuring biology teachers' pedagogical content knowledge (PCK), this article reports on a study about the development of PCK-items for measuring teachers' knowledge of pupils' errors and ways for dealing with them. This study investigated ninth and tenth grade German pupils' (n = 461) drawings in an achievement test about the knee-jerk in biology, which were analysed by using the inductive qualitative analysis of their content. The empirical data were used for the development of the items in the PCK-test. The validation of the items was determined with think-aloud interviews of German secondary school teachers (n = 5). If the item was determined, the reliability was tested by the results of German secondary school biology teachers (n = 65) who took the PCK-test. The results indicated that these items are satisfactorily reliable (Cronbach alpha values ranged from .60 to .65). We suggest a larger sample size and American biology teachers be used in our further studies. The findings of this study about teachers' professional knowledge from the PCK-test could provide new information about the influence of teachers' knowledge on their pupils' understanding of biology and their possible errors in learning biology.

al c Keywords: professional knowledge; biology teachers; pedagogical content knowledge; pupils' errors; pupils' understanding

Introduction

Decades ago, Shulman (1986; 1987) originally defined the professional knowledge of teachers by using several categories. Since then, many research groups have tried to operationalise teachers' professional knowledge (Abell, 2007; Baumert et al., 2010; Park & Oliver, 2008) for the development of measurement instruments. Even now, teachers' professional knowledge, especially the pedagogical content knowledge (PCK), is not easy to define and measure. This article presents a study to develop an instrument for measuring PCK on the basis of empirically analysed pupils' errors. The three steps of the development in this project will be described: (1) analysis and categorisation of the pupils' answers to an achievement test; (2) development of PCK-items according to knowledge about pupils' errors based on the results of step one and determination of the validity of the items; and (3) analysis of the reliability and objectivity of the P. P. developed PCK-test.

Theoretical Background

Professional Knowledge of Science Teachers

In 1987, Shulman proposed seven categories of teachers' professional knowledge. Numerous researchers in the field of science education still use this work to organise research efforts or to design and carry out research studies (e.g., Baumert et al., 2010; Hashweh, 2005; Hill, Ball, & Schilling, 2008; Lee & Luft, 2008; Loughran, Mulhall, & Berry, 2008; Park, Jang, & Chen, 2010; Schmidt et al., 2007; Tatto et al., 2007; van Driel, Beijaard, & Verloop, 2001). For example, Baumert et al. (2010) helped organise their evaluation of teachers' professional knowledge by using three categories: pedagogical content knowledge (PCK), content knowledge (CK) and

pedagogical knowledge (PK). Over the past decade, many research projects have used these three main categories of teachers' professional knowledge to provide a theoretical background. Unlike most of the current studies on mathematics teachers (e.g., Abell, 2007; Baumert et al., 2010; Hill & Ball, 2004; Lipowsky, 2006; Park & Oliver, 2008; Schmidt et al., 2007), the project *ProwiN* (the German acronym for *Professional Knowledge in Science*) (Fischer, Borowski, & Tepner, in press) is a first step to close the gap in research about teachers' professional knowledge in science education (Abell, 2007). Three cooperating German Universities analysed the PCK, the CK and the PK of science teachers based on a theoretical model (Tepner et al., submitted) and the results of researchers-developed test instruments. The study presented here is part of the project *ProwiN* with a focus on one category of the professional knowledge—the PCK of biology teachers.

Theoretical Conceptualisations of the Pedagogical Content Knowledge (PCK)

Most studies on PCK focus on one of the following PCK components: "knowledge about pupils' understanding", "curricular knowledge", "knowledge about instructional strategies and representation" (cf. Kind, 2009; Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008; van Driel, Verloop, & de Vos, 1998). Illustrated here is the development of an instrument for measuring biology teachers' PCK with a special focus on one PCK component—the pupils' errors and their understanding of biology. Of special interest to this study is the *MT21*-study (Schmidt et al., 2007) which analysed two sub-competencies of mathematic teachers' diagnostic competency: recognition of pupils' misconceptions and criteria-guided assessment of pupils' solutions to problems (Schwarz, Wissmach, & Kaiser, 2008).

The PCK in this study (part of the project *ProwiN*) was generally defined according to Shulman (1986) as knowledge of structuring and describing the teaching of the content that would be necessary for preparing the subject matter so that it is comprehensible for pupils (Fischer, Borowski, & Tepner, in press). The most often used PCK components in the literature—knowledge about pupils' understanding, instructional strategies and representations (Park & Oliver, 2008; Schmelzing et al., 2010)—were chosen for the PCK-model of the project (Tepner et al., sub.).

The knowledge about pupils' understanding will be highlighted. This PCK component is defined throughout pupils' errors as well as the knowledge about the thinking of the pupils in solving problems. In addition to the three PCK components of the project (pupils' errors, models and experiments), three knowledge dimensions were defined: declarative, procedural and conditional knowledge (Alexander, Schallert, & Hare, 1991; de Jong & Ferguson-Hessler, 1996; Paris, Lipson, & Wixson, 1983). These knowledge dimensions form the required cognitive aspect of teachers' professional knowledge—here the PCK conceptualisation (Hill, Ball, & Schilling, 2008). Moreover three different biology topics that teachers have to know about—neurobiology, plants and vertebrates—were selected from the curriculum of German schools for the development of the PCK-items. In this article, we use the human reflex arc to demonstrate the idea of developing PCK-items concerning "pupils' errors" and the three different knowledge dimensions (declarative, procedural and conditional).

Pupils' Errors

International Journal of Science Education

In science education, there are many definitions of different terms which have been used for describing pupils' knowledge (Taber, 2009). Sometimes, some terms could be used synonymously and sometimes one and the same term is discussed in different ways. Table 1 shows an overview of the exemplary literature on some of the accepted terms about pupils' knowledge in science that were categorised and combined into two different groups for defining "pupils' errors" in our study.

[Insert Table_1 about here]

Conceptions and prior knowledge. In the literature about pupils' scientifically incorrect conceptions, two main aspects have become apparent—pupils can bring their ideas with them to the classroom from their everyday life or they can develop their ideas during the lesson and/or because of the topic being taught in the lesson. Accordingly, the terms can be classified into two groups. (1) *Conceptions*—defined as a way of pupils' thinking which they established on their own depending on their experience in everyday life but not consistent with teachers' and scientists' ideas—have often been called *misconceptions*, *preconceptions*, *alternative conceptions*, *alternative frameworks* or *children's science* (Treagust, 1988). (2) *Prior knowledge* (Allen, 2005)—described as knowledge before pupils learn something about a topic—has been studied since the 1970s. *Knowledge* is used as a tightly-defined term and *idea*, *belief*, as well as *concept* are sometimes used as synonyms (Taber, 2009). But knowledge differs from conceptions, because knowledge of pupils can be described as a set of conceptions.

Pupils' errors. Concurrent to studies on "conceptions or misconceptions", research in pedagogical psychology explores "*pupils' errors*" (Oser, Hascher, & Spychiger, 1999; Seifried & Wuttke, 2010; Spychiger, 2008). Pupils' errors are the out coming signals of their difficulties or

misunderstandings in lessons. The "errors" are different from the "misconceptions", because misconceptions could be defined as scientifically incorrect knowledge about something which can and should influence pupils' conceptual learning in the lesson (Tanner & Allen, 2005). The meaning between neutral and non-neutral terms can be distinguished, for example, "error" or "knowledge" (Taber, 2009), which indicates the possibility of saying if something is wrong or right according to the current scientific research on the topic. But there are problems within the definition as well—what is wrong and what is right in science (Cokelez, Dumon, & Taber, 2008)? Using these non-neutral—often negatively understood—terms, researchers have to define them clearly for the terms not being misunderstood (Oser, Hascher, & Spychiger, 1999). On the contrary, a neutral term will be, for example, "children's ideas" and "alternative conceptions" (Taber, 2009), which are defined more openly to include most ideas of pupils.

In this study, the term "errors" was used because we were interested in errors of the ninth and tenth grade pupils after their learning process in biology in the ninth grade. The focus was on errors of pupils because these errors could be seen as an indicator of the effectiveness of instruction (Seidel & Shavelson, 2007). Pupils' errors are an important component of teachers' PCK on how to handle errors and to examine what might be the reasons for such errors (Baumert et al., 2010; Hill, Rowan, & Loewenberg Ball, 2005; Lipowsky, 2006; Seifried, & Wuttke, 2010).

Aims

Based on the theoretical background, the following were the research questions and aims of our study:

(1) What are the typical pupils' errors about the reflex arc of the knee-jerk?

- (2) How can items for a PCK-test be developed on the basis of the identified pupils' errors to measure teachers' PCK of pupils' errors and ways for dealing with them?
- (3) Are these PCK-items reliable and valid for assessing the PCK of pupils' errors?

Method

The three parts of the study described here are: pupils' errors analysis, development of the PCKitems as well as the qualitative validation and quantitative analysis of the developed PCK-items.

Participants

An achievement test was given to 461 ninth- and tenth-grade pupils in secondary schools in Germany—204 male pupils (42.7%) and 249 female pupils (52.1%) (25 pupils without gender information) with an average age of M = 15.14 years (33 pupils without age information). All the pupils who took the achievement test (about the reflex arc) had learned about the topic before the test in their ninth grade.

The results of this achievement test were used to develop PCK-items about pupils' errors. To determine the content validity of these items, five German biology teachers—three teachers are teaching at the higher secondary school and two at the general-education secondary school— were interviewed. They had an average age of 43.8 years and an average teaching experience in biology of 13.40 years. Additionally, all the developed items concerning three different biology topics and the PCK about pupils' errors were piloted with a larger sample of 65 biology teachers. They had an average age of 39.54 years and an average teaching experience of 10.50 years (not including the two years of practicum during the apprenticeship to become a teacher). All the

participating teachers—58.5% female and 41.5% male biology teachers from six different federal states of Germany—answered the PCK-test. Of these participating biology teachers, 72.3% taught at the higher secondary schools and 27.7% taught at the general-education secondary schools.

Achievement Test

In one item of the pupils' achievement test, the pupils had to mark on a drawing of a given human body, the pathway of the impulse that travels during a *knee-jerk*—the idea was based on Hammann (2003). The pupils had about five minutes for this item (Figure 1). The expected answer of the pupils would be the pathway of an *impulse* starting at the *quadriceps muscle*, travelling to and through the *white* and *gray matter* of the *spinal cord*, and ending in the muscle again. The pathway from the muscle to the gray matter is the *sensory neuron* and the pathway back to the muscle is the *motor neuron*, which should be labelled as *afferent* and *efferent nerves*.

[Insert fig_1 about here]

Development of the PCK-items

Based on the empirically identified and counted error categories of the pupils' answers to the presented item (see Figure 1), PCK-items for biology teachers were developed in three different knowledge dimensions: declarative knowledge, procedural knowledge and conditional knowledge.

Figure 2 shows an example of three items—about the biological topic "reflex arc"—as part of the PCK-test for analysing pupils' errors with respect to the three different knowledge dimensions:

International Journal of Science Education

Item a) on declarative knowledge; Item b) on conditional knowledge; and Item c) on procedural knowledge. The most frequently counted pupils' errors category (i.e., category 1 in the results of this study, see Table 2) was used for the development of this item.

[Insert fig_2 about here]

At the beginning of the three items (a, b, c), the teachers got pedagogical background information (Baxter & Ledermann, 1999) about the pupils: their grade, the point in time of their achievement test and about how the test looked like. To develop PCK-items which are almost independent of the CK, the pupils' errors were written explicitly next to the drawing. Because some of the pupils' errors can be caused by materials given to them in the lesson, this is one aspect asked by Item 1b). Major aspects of teachers' PCK include not only the knowledge about possible reasons for pupils' errors but also the knowledge about pupils' upcoming errors and the use of these for planning future lessons or knowing how someone might react in special error situations (Seifried & Wuttke, 2010). These aspects are illustrated by the first (Item a) and the third item (Item c), in Figure 2. The idea of the item development is theoretically based on the empirical result that the combination of high diagnostic competency, high teachers' judgement accuracy and high frequency of instructional strategies forms the key for improving instructional quality and therefore pupils' learning outcomes (Helmke & Schrader, 1987).

The coding manual to rate the PCK-items was written on the basis of the literature in science education, the empirically identified error categories and additional results of schoolbook analysis. The answers of the teachers to Item a) were coded according to the empirical results of

the first part of the study by the use of partial credits for exactly correct answers (1), correctly given frequencies in the interval of $\pm 10\%$ of the correct answer (0.5) and other answers (0). For coding of answers to Item b), the number of meaningful named reasons was counted. Results of the document analysis (schoolbooks; school material) as well as literature (Schumacher, 2008; Seifried & Wuttke, 2010) formed the basis of this coding manual. The idea behind this item was that the more meaningful reasons the teachers name, the more the teachers might be able to consider pupils' errors in lesson planning. This coding manual was developed concerning the qualitative document analysis (schoolbooks; lesson materials; everyday life material like advertisements). Of course, if a teacher did not name a meaningful reason in the list of examples of reasons in the coding manual, this was counted with one point as well.

The possible reactions of teachers to pupils' errors in the described situation in the biology lessons (Item c) were coded with one point per each named reaction. Therefore, the listed possible reactions of the teachers are summarised in four main categories based on theoretical background: (1) error is used as orientation for further lessons; (2) error is used for other pupils so that they could learn from others errors' (analysis of errors); (3) error corrected by the pupil her-/himself so that she/he can learn from her/his errors; and (4) teachers being flexible and describe the misunderstood topic another way around during the lesson (Jüttner & Neuhaus, 2010; Caspary, 2008; Oser & Spychiger, 2005; Taber, 2009).

All the items concerning "neurobiology" and all PCK components (pupils' errors, models, experiments) were validated in an additional study. Similar items concerning the other biology topics (plants and vertebrates) were piloted. For the component "pupils' errors" there are always three PCK-items per topic according to the knowledge dimensions. Additionally, for the PCK component "model", always one item per topic—about the reasons of pupils' errors related to the

model and its use—could be added to the PCK-items about pupils' errors. So all in all, 12 PCKitems concerning the knowledge about pupils' errors were created.

Data Analysis Methods

The categories of the pupils' errors were identified using the inductive qualitative analysis of content for 10% of the pupils' answers (Hsieh & Shannon, 2005; Mayring, 2000). The inductively generated categories were described in the coding manual with which multiple counting was possible. After the coding of pupils' answers to the shown item by two independent raters, the relative frequency for each category was computed by using the statistical software SPSS 19.0.0 (2011). To analyse the interrater reliability for 10% of the sample (n = 46) the AC1 statistic (Gwet, 2002; 2008) was calculated in addition to the commonly used Cohen's kappa value (Landis & Koch, 1977; Shoukri, 2004). This was caused by the paradoxical kappa values because of the seldom coded categories (Gwet, 2002). Additionally, ten typical German schoolbooks were analysed to identify information about possible reasons for the categorised pupils' errors.

The content validity of the developed PCK-items had been addressed throughout the process of the development—the empirical data about pupils' errors were used for the creation of the items of the same biological topic. On the other hand, congruent think-aloud interviews were used which implicates that the teachers had to communicate aloud their thoughts and simultaneously write them down on the paper-and-pencil test (Ericsson & Simon, 1993; Leighton, 2009). After the verbatim transcription of the interviews, the two independent raters coded the interview transcripts by using a developed coding manual based on the theoretical model (Tepner et al.,

sub.) underlying the items. Therefore, the intraclass correlation coefficient (ICC) of the two independent raters was calculated (Field, 2009; Shrout & Fleiss, 1970).

The reliability of the PCK-test was analysed by using the Cronbach's alpha coefficient (Field, 2009). The objectivity was tested by using the intraclass correlation coefficient (ICC) concerning the coding of teachers' answers to the open-ended PCK-test items (Field, 2009).

Results

Categories of Pupils' Errors (based on pupils' drawings about the reflex arc)

Seven main categories of pupils' errors were identified and counted in the pupils' (n = 461)answers. Table 2 shows an overview of the frequencies of the seven categories. Some of the most often identified error categories are discussed with the following three examples of pupils' rf answers (see Figures 3, 4 and 5).

[Insert Table_2 about here]

[Insert fig_3 about here]

Figure 3 shows an example of a pupils' answer for four different categories of errors. The most frequent category 1-the drawn impulse reaches the brain (63%)-was identified in this example. Additionally, two aspects of the second category were identified: beginning and/or ending of the nerve fibres are not located at the muscle, but at bones (42%); and beginning and/or ending of the nerve fibres are not located at the muscle, but in the heel or at the tibia

(14%). Also this example could be used to illustrate the fourth category: *the clearing up centre of the nerves is located in the brain* (26%). The final category identified in this example was category 5: *nerves are located in the body (stomach) and not in the spinal cord* (17%). In the drawing, a dot was used by the pupil to represent the change-over point of the pathway in the brain.

[Insert fig_4 about here]

In contrast to the example in Figure 3, the most frequently identified category 1 was not found in the pupil's answer in Figure 4. In the drawing, the pupil did not connect the brain to the reflex arc of the knee-jerk. Here, category 2b was identified: *beginning and/or ending of the nerve fibres are not located at the muscle, but at bones* (42%). Moreover, there was *no direct connection between the two nerve fibres so that the reflex arc is not closed*, that is an example for category 6 (15%).

[Insert fig_5 about here]

To show examples for every category of pupils' errors, here is the final sample of another pupil's answer (see Figure 5), where the remaining categories of errors were identified. First of all, the most frequent category 1 was identified. Furthermore, *the starting- and ending-point of the nerve fibres is in the front of the patella* (category 2a; 47%) as well as category 3: *only one way for the impulse illustrated by one single nerve* (34%) were identified. At last, the least frequent category 7 could be identified: *nerves are not named according to their function, but to their absolute location (brain = afferent nerve; muscle = efferent nerve)* (7%).

For the interrater reliability of this coding, the Cohen's kappa (κ) for all the categories was .67 ($p \le .001$, n = 462 categorisations). Because of the distribution of the data, the kappa coefficients for five categories of errors were very low (e.g., for category 2a, $\kappa = .35$), negative or even not computable; therefore, the AC1 statistic is described additionally for each category in Table 3 (Gwet, 2008; Schori, Kersten, & Abderhalden, 2006).

[Insert Table_3 about here]

The AC1 statistic (Gwet, 2008) showed values from moderate (AC1 (2a) = .50) to almost perfect (e.g., AC1 (3) = 1) interrater agreement between the two independent raters concerning all the seven pupils' error categories (see Table 3).

Validation of the Developed PCK-items

The results of the analysis of the coding of the teachers' think-aloud interview transcripts could demonstrate how often the teachers thought about or used a knowledge dimension to answer the given PCK-items (see Table 4).

[Insert Table_4 about here]

For example, to answer the third item dealing with conditional knowledge on pupils' errors (PEcond), the five teachers used, on the average, in 90.1% of their thoughts the conditional knowledge on pupils' errors. The intention of this item was to get information about teachers' conditional knowledge about pupils' errors. Additionally, while answering this item, they needed

 or used, on the average, 33% of the declarative knowledge about pupils' errors as well as 16.7% of declarative content knowledge.

The interrater reliability of the two independent raters for the coding of the think-aloud interview transcripts was found to be significantly high (*ICC* uniust = .99; $F_{439,439} = 143.15$; p < .001).

More Quality Criteria of the PCK-items.

For the internal consistency of the PCK-test scale "declarative knowledge" and the PCK component "pupils' errors", a very low Cronbach's alpha value ($\alpha = .06$; 3 items; n = 65) was found. For the two other knowledge dimensions, the Cronbach's alpha values was $\alpha = .60$ and $\alpha = .67$ (n = 65) (see Table 5).

[Insert Table_5 about here]

Finally, the intraclass correlation coefficient (ICC) of the two independent raters for the openended PCK-items was found to be statistically significant (*ICC* _(unjust) = .80; $F_{588,588}$ = 5.10; p < .001).

Discussion and Conclusions

The seven categories of pupils' errors—identified and analysed in their drawings in the achievement test item about the pathway of the impulse of the reflex arc—could be used to analyse pupils' errors, but only for this specific topic. In mathematics, ideas of general error categories exist (e.g., Ashlock, 2005), but not yet in biology. The seven error categories of ninth

International Journal of Science Education

and tenth grade German pupils reported in this article might be summed up as three main categories which might be transferred to other biological topics as well: *wrong locations of the structure*; *errors concerning the function* and *wrongly used terms and definitions*. The results of this study showed that most of the analysed pupils' error categories concerning the knee-jerk were usable for the categorisation of pupils' answers in this topic and that the interrater agreement (AC1) of the categories ranged from moderate to almost perfect (Gwet, 2002; Landis & Koch, 1977).

All in all, this study was able to show one way of developing PCK-items based on categorisation of pupils' errors which might enable researchers to develop their topic-specific, practical-orientated and empirically based PCK-items based on this idea. The quality of the developed PCK-items was validated by qualitative think-aloud interviews and the quantitative pilot phase. The results of these analyses demonstrated the development of a valid, reliable and objective PCK-test instrument for testing teachers' PCK concerning pupils' errors in biology.

The developed items in the PCK-test have curricular content validity because they were developed after curricular reviews and many meetings with experienced teachers. Additionally, the empirical data of the pupils' errors gave more content validity. Finally, the think-aloud interviews demonstrated that teachers mostly thought of or used the kind of knowledge (declarative, procedural or conditional) for answering the item which was intended to be measured by the respective item. The ICC of the interview transcript coding showed a satisfactory interrater agreement between the two independent raters (Field, 2009). In the end of all the interviews, four follow-up questions about the items were also asked. One of these questions asked the teacher if their answers to the items might represent their personal way of enactment in classroom situations or their way of thinking concerning example models in biology

lessons. This could be seen as face validity, because all the interviewed teachers said something similar to what Teacher K said: "My answers represent what I strive to do".

The internal consistency of the PCK-test (its Cronbach's alpha values) showed satisfactory results except for the declarative knowledge of pupils' errors—about teachers' prediction of how often the given pupils' error might arise after the learning process. One reason might be that teachers often wrongly predict their pupils' learning outcomes and because they have over-expectations of their pupils' learning outcomes (Helmke & Schrader, 1987; Hosenfeld, Helmke & Schrader, 2002).

The objectivity of the PCK-items was also confirmed with an acceptable intraclass correlation coefficient (ICC).

Due to the small sample size of this study for more detailed statistical analysis (e.g., factor analysis), further studies with samples of about 200 teachers will be conducted to test if all the developed items conform to the three dimensional theoretical model of the current project *ProwiN* (Tepner et al., sub.). Additionally, the items were translated into English and will be given to American teachers in our further studies. The results of the future think-aloud interviews might show the possible use and transferability of this study's development ideas to other countries and cultures.

Teachers' knowledge about pupils' possible errors—which may arise in a lesson or about possible reasons for these errors—is important for teachers because this knowledge has different advantages for teachers. On the one hand, it can be used as an orientation during the planning process of lessons (Schumacher, 2008; Seifried &Wuttke, 2010); and meaningful arrangement of materials, figures and models can also result. On the other hand, the information can be used for

the training of pupils' error analysis; pupils have the possibility to learn the so-called *negative knowledge* (Oser, Hascher, & Spychiger, 1999). In the analysis of errors made by anonymous pupils, the pupils have to change their point of view to that of a corrector. This change is the best opportunity for learning from errors (Schumacher, 2008). Therefore, empirically analysed and categorised pupils' errors, as reported in this article, are important for teachers. In addition, these results can be used for test items construction in science education research as well as in teacher education.

To conclude, the method of developing the PCK-items demonstrated in our study was able to test teachers' knowledge by its connectedness to praxis, as well as to relate pupils' performance in achievement tests directly to teachers' knowledge about their pupils' errors. So it will be possible to construct tests for pupils and teachers for analysing the correlation between pupils' knowledge and teachers' PCK concerning pupils' understanding.

As for research, the aim will be to analyse the PCK of teachers and to see if their knowledge correlates with how they prepare and teach the subject matter in different biological topics as well as with their job experience. So far, mathematical pre-service teachers' experience has been analysed in longitudinal studies (e.g. Tatto et al., 2008). Additionally, there exist a few qualitative studies with a smaller sample size evaluating the development of teachers' professional knowledge (Henze & van Driel, 2009; Friedrichsen, Abell, Pareja, Brown, Lankford, &Volkmann, 2007). In research there are different results concerning the influence of teachers' experience is experience. Friedrichsen et al. (2007) found that the development of science teachers' pedagogical knowledge was supported by experience but the PCK development was not. Further research concerning the analysis and conceptualisation of general pedagogical knowledge is

URL: http://mc.manuscriptcentral.com/tsed Email: editor_ijse@hotmail.co.uk

 needed (Friedrichsen et al., 2007; Blömeke, Felbrich, & Müller, 2008). As well, biology teachers' professional knowledge development (longitudinal) and the relationship between the developments of the different components (PCK, CK, PK) could be focused on in the future to close the gap for biology teachers' research (Cohen & Yarden, 2009; Abell, 2007).

References

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105-1149). New Jersey: Lawrence Erlbaum Associates.
- Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61 (3), 315-343.
- Allen, M. (2005). Encouraging pupil error may promote better understanding of a scientific concept. *School Science Review*, 86, 79–85.
- Allen, M. (2010). Learner error, affectual stimulation, and conceptual change. *Journal of Research in Science Teaching*, 47 (2), 151-173.
- Ashlock, R. B. (2005). Error patterns in computation: Using error patterns to improve instruction (9th Edition). New-York: Macmillan.
- Bainbridge, R. (1981). To err is human: Towards a more positive approach to young children's mistakes in arithmetic. *Mathematics in School*, 10(5), 10–13.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers' mathematical knowledge, cognitive cctivation 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.

Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des

erziehungswissenschaftlichenWissens angehender Lehrkräfte [Assessment of student teachers' pedagogical knowledge]. In S. Blömeke, G. Kaiser & R. Lehmann (Eds.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 171–194). Münster, Germany: Waxmann.

Bouvier, A.(1987). The right to make mistakes. For the Learning of Mathematics, 7(3), 17–25.

- Caleon, I. S., & Subramaniam, R. (2010). Do pupils know what they know and what they don't know? Using a Four-Tier diagnostic test to assess the nature of pupils' alternative conceptions. *Research Science Education*, 40, 313-337.
- Chi, M. T. H., Slotta, J. D., & DeLeeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.
- Cokelez, A., Dumon, A., & Taber, K. S. (2008). Upper secondary French pupils, chemical transformations and the "Register of models". *International Journal of Science Education*, 30 (6), 807-836.
- Cohen, R. & Yarden, A. (2009). Experienced Junior-High-School teachers' PCK in light of a curriculum change: "The cell is to be studied longitudinal". *Research of Science Education*, 39, 131-155.

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105 - 113.

de Laughter, J. E., & Stein, S. (1998). Preconceptions about earth science among pupils in an introductory course. *Eos*, 79, 429-434.

Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629-646). New York, London: Routledge.

Driver, R.P. (1981). Pupils' alternative frameworks in science. *European Journal of Science Education*, 3(1), 93–101.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: The MIT Press. Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.

- Fischer, H.E., Borowski, A., & Tepner, O. (in press). Professional knowledge of science teachers;In B. Fraser; K. Tobin, & C. McRobbie (Eds.). Second International Handbook of Science Education, New York: Springer.
- Friedrichsen, P., Abell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46, 357-383.
- Gilbert, J., Osborne, R., & Fensham, P. (1982). Children's science and its consequences for teaching. *Science Education*, 66, 623-33.
- Gwet, K. L. (2008). *Interrater Reliability*. Wiley Encyclopedia of Clinical Trials, Copyright 2008 John Wiley & Sons, Inc.

Gwet, K. L. (2002). Kappa Statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods For Inter-Rater Reliability Assessment*, 1, 1-5. Retrieved from http://agreestat.com/research_papers/kappa_statistic_is_not_satisfactory.pdf [Date accessed: 25 May. 2011].

Hammann, M. (2003). Aus Fehlern lernen. [Learning by errors], Unterricht Biologie, 287, 31-35.

- Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of PCK. *Teachers and Teaching: Theory and Practice*, 11, 273–292.
- Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3, 91-98
- Helm, H. (1980). Misconceptions in physics amongst South African pupils. *Physics Education*, 15, 92-7.
- Henze, I., Driel van, J. H., & Verloop, N. (2008). Development of experienced science teachers' pedagogical content knowledge of models of the solar system and the universe. *International Journal of Science Education*, 30(10), 1321-1342.
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35 (5), 330-351.
- Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge:Conceptualizing and measuring teachers' topic-specific knowledge of pupils. *Journal for Reserach in Mathematics Education*, 39, 372-400.

Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichtsund lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Lehr-Lern-Studie SALVE. [Diagnotic competency: teaching and learning relevant to student characteristics and the teachers' prediction in the teaching-learning study SALVE], In M. Prenzel & J. Doll (Eds.),*Bildungsqualität von Schule: Schulische und auβerschulische*

Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. Zeitschrift für Pädagogik, 45. Beiheft (pp. 65-82). Beltz: Weinheim.

- Houssart, J., & Weller, B. (1999). Identifying and dealing with misconceptions and errors in primary mathematics: Student teachers record their experiences. *Mathematics Education Review*, 11, 46-58.
- Hsieh, H.-F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Jüttner, M., & Neuhaus, B. (2011). Vom Schülerfehler zum PCK Test. Ansätze zur Entwicklung von PCK – Items. [A PCK-test based on students' errors. Rudiments for developing PCK items.] In U. Harms & I. Mackensen-Friedrichs (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik. Band 4* (pp. 27-39).Studien Verlag: Innsbruck.
- Jüttner, M. & Neuhaus, B. (2010). Using empirically analyzed pupils' errors to develop a PCK-Test. In M.F. Taşar & G. Çakmakcı (Eds.), *Contemporary science education research: preservice and inservice teacher education* (pp. 331-340). Ankara, Turkey: Pegem Akademi.
- Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education*, 45(29), 169-204.
- Krüger, D. (2007). Die Conceptuale Change-Theorie. [The conceptual change theory]. In D.Krüger & H. Vogt (Eds.), *Handbuch der Theorien in der biologiedidaktischen Forschung* (pp. 81-92). Berlin: Springer Verlag.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159–174.
- Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, 30, 1343-1363.

- Leighton, J. P.(2009). Two Types of Think Aloud Interviews for Educational Measurement:Protocol and Verbal Analysis. *Paper presented at the National Council on Measurement in Education (NCME)*, CA.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. [It depends on the teacher. Empirical evidences of the connection between teachers' competence, teachers' acting and the learning of the pupils], *Zeitschrift für Pädagogik*, 52 (51), 47-70.
- Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education*, 30 (10), 1301-1320.
- Magnusson, S., & Krajcik, J. & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In: Gess-Newsome & Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-131). Dordrecht. Kluwer Academic Publishers.
- Mayring, P. (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, North America, 1. Available at: http://www.qualitativeresearch.net/index.php/fqs/article/view/1089/2385. [Date accessed: 26 Oct. 2010].
- Nebah, S. J. (2002). Do people learn from mistakes? A pedagogical study and analysis of autobiographies and interviews. Dissertation am Departement Erziehungswissenschaften der Universität Fribourg.
- Novak, G.S. (1977). Representations of knowledge in a program for solving physics problems. Proceedings of the Fifth International Joint Conference on Artificial Intelligence. Cambridge, MA: MIT Press.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie der "negativen" Wissens. [Learning out of errors. About the psychology of the "negative"

knowledge]. In W. Althof (Ed.), *Fehlerwelten: vom Fehlermachen und Lernen aus Fehlern* (pp. 11-43). Opladen: Leske + Budrich.

- Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur*.[Learning is painful. About the theory of the negative knowledge and about the practice of the culture of errors]. Weinheim: Beltz.
- Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. Contemporary *Educational Psychology*, 8, 293-316.
- Park, S., Jang, J.-Y., & Chen, Y.-C. (2010). Is Pedagogical Content Knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, 41, 245–260. doi: 10.1007/s11165-009-9163-8
- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a conceptual tool to unterstand teachers as professionals. *Research in Science Education*, 38, 261-284.
- Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Measuring declarative and reflective components of biology teachers' pedagogical content knowledge. In M. F. Tasar & G. Cakmakci (Eds.), *Contemporary science education research: Teaching* (pp. 71-77).
 Ankara, Turkey: Pegem Akademi.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L. et al. (2007). *The Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries. Mathematics Teaching in the 21st Century (MT21)* (Rep. No. MSU Center for Research in Mathematics and Science Education).
- Schori, E., Kersten, B. & Abderhalden, C. (2006). Wie reliabel ist der Fragebogen zur Analyse der Pflegedokumentation IzEP-Dok© im Instrument zur Erfassung von Pflegesystemen IzEP©? [How Reliable is the Questionnaire for the Analysis of Nursing Documentation

IzEP-Dok[©] as part of the Measurement Instrument for Collecting Data of the Care Systems IzEP[©]?], *Printernet Angewandte Pflegeforschung*, 8, 1-9

- Schumacher, R. (2008). Der produktive Umgang mit Fehlern. Fehler als Lerngelegenheit und Orientierungshilfe. [The productive handling with errors. Errors as learning opportunity and for orientation]. In R.Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp. 7-11). Freiburg: Herder.
- Schwarz, B., Wissmach, B., & Kaiser, G. (2008). "Last curves not quite correct": diagnostic competences of future teachers with regard to modelling and graphical representations. *ZDM The International Journal on Mathematics Education*, 40(5), 777-790.
- Seifried, J., & Wuttke, E. (2010). Student errors: how teachers diagnose them and how they respond to them. *Empirical Research in Vocational Education and Training (ERVET)*, 2(2), 147-162.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Education Research*, 77, 454-499.

Shoukri, M. M. (2004). Measures of interobserver agreement. Boca Raton: CRC PR Inc.

- Shulman, L. (1986). Those who understand. Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the New Reform. *Havard Educational Review*, 57, 1-22.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86 (2), 420–428.
- Spychiger, M. (2008). Ein offenes Spiel. Lernen aus Fehlern und Entwicklung von Fehlerkutur. [An open play. Learning by errors and the development of the culture of errors]. In R.

Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp. 25-48). Freiburg: Herder.

- Taber, K. S. (2009). *Progressing science education: Constructing the scientific research programme into the contingent nature of learning science*. Dordrecht: Springer.
- Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: from a scholarly approach to teaching to the scholarship of teaching. *Cell Biology Education*, 4, 1-6.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

- Tepner, O., Borowski, A., Fischer, H.E., Jüttner, M., Kirschner, S., Leutner, D. et al. (submitted). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. [Theoretical model for the developement of testitems to collect data about professional knowledge of science teachers], Zeitschrift der Naturwissenschaften.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10 (2), 159-169.
- Tsai, C.-C., & Chou, C. (2002). Diagnosing pupils' alternative conceptions in science. *Journal of Computer Assisted Learning*, 18, 157–165.

Tsamir, P., Rasslan, S., & Dreyfus, T. (2006). Prospective teachers' reactions to Right-or-Wrong tasks: The case of derivatives of absolute value functions. *Journal of Mathematical Behavior*, 25, 240–251.

van Driel, J. H., Verloop, N., & Vos, W. d. (1998). Developing science teachers' Pedagogical Content Knowledge. *Journal of Research in Science Teaching*, 35, 673-695. van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. Journal of Research in Science

 . Ministry of Education and Research; it is a cot,

 .a in education' (01JH0904).

 Teaching, 38, 137-158.

* the presented project is funded by the Federal Ministry of Education and Research; it is a cooperation project embedded in the framework program of 'empirical research in education' (01JH0904).

Table 1

A List of Selected Terms Used in Science Education concerning Pupils' Knowledge

Term	Exemplary Literature	Deleted (mis)essentian	
conceptions or misconceptions,	Allen, 2010; Chi, Slotta, & de Leewu, 1994;		
	Duit, Treagust, & Widodo, 2008; Helm, 1980;		
	Krüger 2007 · Treagust 1988·		
	1. uger, 2007, 1. eugaol, 1900,		
alternative conception	Caleon & Subramaniam, 2010; Taber, 2009;		
	Tsai & Chou. 2002		
alternative frameworks	Driver, 1981		
preconceptions	Novak, 1977; de Laughter & Stein, 1998		
mistakes	Brainbridge, 1981; Bouvier, 1987; Nebah,		
	2002		
errors	Ashlock, 2005; Houssart & Weller, 1999;		
	Tsamir, Rasslan, & Dreyfus, 2006; Seifried &		
	Wuttke 2010		
		Deleted: C	
children's science	Gilbert et al., 1982	Deleted: '	
<u>children's science</u>	Gilbert et al., 1982		

Below the human contour of a child is drawn. Structures of the brain, the spinal cord and

Please draw the pathway of the knee jerk reflex in the child's body in the above drawing

Figure 1.

The item of the pupils' achievement test for analysing pupils' errors about the reflex arc after the

learning process in the ninth and tenth grade (n = 461).

112x78mm (300 x 300 DPI)

important muscles, bones and fibres are given as well.

and label the efferent and afferent nerves.





Example of three items as a part of the PCK-test for measuring teachers' knowledge of pupils' errors and ways for dealing with them. 234x317mm (300 x 300 DPI)

ummary of Seven Identified Categories of Typical Students' Errors (n = <u>461</u>) al erk (<u>cf. Jüttner & Neuhaus, 2011</u>)	out Knee-	Deleted: authors, in pr
Identified categories of pupils' errors	frequency	Deleted: student
1.) the drawn impulse reaches the brain	<u>63</u> %	
2.) beginning and/or ending of the nerve fibres are not located at the muscle	,	
but		
a. in front of the patella	<u>47%</u>	
b. at bones	4 <u>2</u> %	
c. in the heel or at the tibia	<u>14</u> %	
d. on the tendon	<u>12</u> %	
3.) there is only one way for the impulse illustrated by one single nerve	<u>34</u> %	
4.) the clearing up centre of the nerves is located in the brain	<u>26%</u>	
5.) nerves are located in the body (stomach) and not in the spinal chord	17%	
6.) nerves do not come together in the spinal cord/the reflex arc is not closed	l <u>15</u> %	
7.) nerves are not named according to their function, but to their absolute	<u>7</u> %	
location (brain = afferent nerve; muscle = efferent nerve)		



Figure 3. Example of a pupil's answer to the task on the reflex arc (Figure 1). Category 1 as well as category 2b, 2c, 4 and 5 are identified here. 136x133mm (600 x 600 DPI)







Figure 4. Another pupil's answer to the item seen in Figure 1. Category 2b and 6 could be identified here. 126x133mm (600 x 600 DPI)



Figure 5. The sample of a pupil's answer to the item seen in Figure 1. Category 1, 2a, 3 and 7 could be identified here. 162x135mm (600 x 600 DPI)

Table 3

<u>Results and Interpretation of the Interrater Reliability (AC1 Statistic) (Gwet, 2008; Landis &</u> <u>Koch, 1977; Schori, Kersten, & Abderhalden, 2006)</u>

AC1 statistic	Strength of agreement	AC1 for the coding (categories)
< 0.00	Poor	
0.00 - 0.20	<u>Slight</u>	
0.21 - 0.40	<u>Fair</u>	
0.41-0.60	Moderate	<u>AC1 (2a) = 0.50</u>
<u>0.61 – 0.80</u>	Substantial	<u>AC1 (2d) = 0.76; AC1(4) = 0.79; AC1(6) = 0.70</u>
0.81 - 1.00	Almost perfect	AC1(1) = 0.87; AC1(2b) = 0.82; AC1(2c) = 0.90;
		AC1(3) = 1; AC1(5) = 1; AC1(7) = 0.92

Table 4

<u>Means and Standard Deviations for the First Coder's Coded Categories of Teachers' (n = 5)</u> <u>Think-aloud Interview Transcripts Concerning Developed PCK-items (cf. Figure 2)</u>

<u>Item</u>	<u>N</u>	PCK about pupils' errors (PE)		r <u>s (PE)</u>	<u>CK</u>	<u>CK</u>
		declarative	procedural	<u>conditional</u>	(declarative)	(diagnose of
		knowledge	knowledge	knowledge		pupils' errors)
PEdec	<u>5</u>	M = 63%	=	<u><i>M</i></u> = 41.3%	<u></u>	<u><i>M</i></u> = 20 %
		<u>(SD = 22%)</u>		<u>(SD = 10.2%)</u>		(SD = 0)
PEproc	<u>5</u>	<u><i>M</i> = 50%</u>	<u><i>M</i> = 70.8%</u>	<u>M = 100%</u>	<u><i>M</i> = 33.3%</u>	
		(SD = 0)	<u>(SD = 21.7%)</u>	(SD = 0)	<u>(SD = 16.7%)</u>	
PEcond	<u>5</u>	<u><i>M</i> = 33%</u>		M = 90.1%	<u><i>M</i> = 16.7%</u>	
		(SD = 0)		<u>(SD = 13.2%)</u>	(SD = 0)	

 Formatted: Left: 70.85 pt, Right:
70.85 pt, Top: 70.85 pt, Bottom:
56.7 pt, Width: 595.3 pt, Height:
841.9 pt, Header distance from edge:
35.4 pt, Footer distance from edge:
35.4 pt

Deleted: 4

Deleted: c Deleted: d Deleted: Deleted: d Deleted: f Deleted: n Deleted: e Deleted: a Deleted: t Deleted: m Deleted: Deleted: 0 Deleted: 0 Deleted: 0 Deleted: all knowledge dimensions Deleted: pupils' errors Deleted: 9 **Deleted:** $\alpha = 0.71$ Deleted: 0

Table 5_{--}

Reliability (<u>Cronbach's alpha</u>) of the <u>Different PCK-scales</u> <u>Dealing with the Facet 'Pupils'</u> <u>Errors' According to the Theoretical Model (n = 65).</u>

knowledge dimension	facet	items	<u>C</u> ronbach <u>'</u> s alpha		De
declarative knowledge	pupils' errors	<u>3</u>	<u>a = .06</u>	- milit	De
procedural knowledge	nupils' errors	3	a = 60		De
procedurar knowledge		<u>-</u>	<u>« – •••</u>		De
conditional knowledge	pupils' errors	6	$\alpha = 67$		De
۲		¥	x		De
all knowledge dimensions	pupils' errors	12	a = 65		De
	pupils citors	12	<u>« – •••</u>	- "","," - "",",","	De
					De
					De
				, ,	De
					De