



**HAL**  
open science

## 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 Aug 2012

**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.



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

Journal:	<i>International Journal of Science Education</i>
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.

1  
2  
3 Keywords: *professional knowledge; biology teachers; pedagogical content knowledge; pupils'*  
4  
5  
6 *errors; pupils' understanding*  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

## 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 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

1  
2  
3 pedagogical knowledge (PK). Over the past decade, many research projects have used these three  
4  
5 main categories of teachers' professional knowledge to provide a theoretical background. Unlike  
6  
7 most of the current studies on mathematics teachers (e.g., Abell, 2007; Baumert et al., 2010; Hill  
8  
9 & Ball, 2004; Lipowsky, 2006; Park & Oliver, 2008; Schmidt et al., 2007), the project *ProwiN*  
10  
11 (the German acronym for *Professional Knowledge in Science*) (Fischer, Borowski, & Tepner, in  
12  
13 press) is a first step to close the gap in research about teachers' professional knowledge in science  
14  
15 education (Abell, 2007). Three cooperating German Universities analysed the PCK, the CK and  
16  
17 the PK of science teachers based on a theoretical model (Tepner et al., submitted) and the results  
18  
19 of researchers-developed test instruments. The study presented here is part of the project *ProwiN*  
20  
21 with a focus on one category of the professional knowledge—the PCK of biology teachers.  
22  
23  
24  
25  
26  
27  
28  
29

### 30 *Theoretical Conceptualisations of the Pedagogical Content Knowledge (PCK)*

31  
32  
33 Most studies on PCK focus on one of the following PCK components: “knowledge about pupils’  
34  
35 understanding”, “curricular knowledge”, “knowledge about instructional strategies and  
36  
37 representation” (cf. Kind, 2009; Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008; van  
38  
39 Driel, Verloop, & de Vos, 1998). Illustrated here is the development of an instrument for  
40  
41 measuring biology teachers' PCK with a special focus on one PCK component—the pupils’  
42  
43 errors and their understanding of biology. Of special interest to this study is the *MT21*-study  
44  
45 (Schmidt et al., 2007) which analysed two sub-competencies of mathematic teachers' diagnostic  
46  
47 competency: recognition of pupils' misconceptions and criteria-guided assessment of pupils'  
48  
49 solutions to problems (Schwarz, Wissmach, & Kaiser, 2008).  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 The PCK in this study (part of the project *ProwiN*) was generally defined according to Shulman  
4 (1986) as knowledge of structuring and describing the teaching of the content that would be  
5 necessary for preparing the subject matter so that it is comprehensible for pupils (Fischer,  
6 Borowski, & Tepner, in press). The most often used PCK components in the literature—  
7 knowledge about pupils' understanding, instructional strategies and representations (Park &  
8 Oliver, 2008; Schmelzing et al., 2010)—were chosen for the PCK-model of the project (Tepner et  
9 al., sub.).  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20  
21  
22 The knowledge about pupils' understanding will be highlighted. This PCK component is defined  
23 throughout pupils' errors as well as the knowledge about the thinking of the pupils in solving  
24 problems. In addition to the three PCK components of the project (pupils' errors, models and  
25 experiments), three knowledge dimensions were defined: declarative, procedural and conditional  
26 knowledge (Alexander, Schallert, & Hare, 1991; de Jong & Ferguson-Hessler, 1996; Paris,  
27 Lipson, & Wixson, 1983). These knowledge dimensions form the required cognitive aspect of  
28 teachers' professional knowledge—here the PCK conceptualisation (Hill, Ball, & Schilling,  
29 2008). Moreover three different biology topics that teachers have to know about—neurobiology,  
30 plants and vertebrates—were selected from the curriculum of German schools for the  
31 development of the PCK-items. In this article, we use the human reflex arc to demonstrate the  
32 idea of developing PCK-items concerning “pupils' errors” and the three different knowledge  
33 dimensions (declarative, procedural and conditional).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

#### 51 52 53 54 *Pupils' Errors* 55 56 57 58 59 60

1  
2  
3 In science education, there are many definitions of different terms which have been used for  
4 describing pupils' knowledge (Taber, 2009). Sometimes, some terms could be used  
5 synonymously and sometimes one and the same term is discussed in different ways. Table 1  
6 shows an overview of the exemplary literature on some of the accepted terms about pupils'  
7 knowledge in science that were categorised and combined into two different groups for defining  
8 "pupils' errors" in our study.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20 [Insert Table\_1 about here]  
21  
22  
23

24 *Conceptions and prior knowledge.* In the literature about pupils' scientifically incorrect  
25 conceptions, two main aspects have become apparent—pupils can bring their ideas with them to  
26 the classroom from their everyday life or they can develop their ideas during the lesson and/or  
27 because of the topic being taught in the lesson. Accordingly, the terms can be classified into two  
28 groups. (1) *Conceptions*—defined as a way of pupils' thinking which they established on their  
29 own depending on their experience in everyday life but not consistent with teachers' and  
30 scientists' ideas—have often been called *misconceptions*, *preconceptions*, *alternative*  
31 *conceptions*, *alternative frameworks* or *children's science* (Treagust, 1988). (2) *Prior knowledge*  
32 (Allen, 2005)—described as knowledge before pupils learn something about a topic—has been  
33 studied since the 1970s. *Knowledge* is used as a tightly-defined term and *idea*, *belief*, as well as  
34 *concept* are sometimes used as synonyms (Taber, 2009). But knowledge differs from  
35 conceptions, because knowledge of pupils can be described as a set of conceptions.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

52  
53 *Pupils' errors.* Concurrent to studies on "conceptions or misconceptions", research in  
54 pedagogical psychology explores "*pupils' errors*" (Oser, Hascher, & Spychiger, 1999; Seifried &  
55 Wuttke, 2010; Spychiger, 2008). Pupils' errors are the out coming signals of their difficulties or  
56  
57  
58  
59  
60



1  
2  
3 misunderstandings in lessons. The “errors” are different from the “misconceptions”, because  
4  
5 misconceptions could be defined as scientifically incorrect knowledge about something which  
6  
7 can and should influence pupils’ conceptual learning in the lesson (Tanner & Allen, 2005). The  
8  
9 meaning between neutral and non-neutral terms can be distinguished, for example, “error” or  
10  
11 “knowledge” (Taber, 2009), which indicates the possibility of saying if something is wrong or  
12  
13 right according to the current scientific research on the topic. But there are problems within the  
14  
15 definition as well—what is wrong and what is right in science (Cokelez, Dumon, & Taber,  
16  
17 2008)? Using these non-neutral—often negatively understood—terms, researchers have to define  
18  
19 them clearly for the terms not being misunderstood (Oser, Hascher, & Spychiger, 1999). On the  
20  
21 contrary, a neutral term will be, for example, “children’s ideas” and “alternative conceptions”  
22  
23 (Taber, 2009), which are defined more openly to include most ideas of pupils.  
24  
25  
26  
27  
28  
29  
30  
31

32 In this study, the term “errors” was used because we were interested in errors of the ninth and  
33  
34 tenth grade pupils after their learning process in biology in the ninth grade. The focus was on  
35  
36 errors of pupils because these errors could be seen as an indicator of the effectiveness of  
37  
38 instruction (Seidel & Shavelson, 2007). Pupils’ errors are an important component of teachers’  
39  
40 PCK on how to handle errors and to examine what might be the reasons for such errors (Baumert  
41  
42 et al., 2010; Hill, Rowan, & Loewenberg Ball, 2005; Lipowsky, 2006; Seifried, & Wuttke, 2010).  
43  
44  
45  
46  
47

#### 48 *Aims*

49  
50 Based on the theoretical background, the following were the research questions and aims of our  
51  
52 study:  
53  
54

- 55 (1) What are the typical pupils’ errors about the reflex arc of the knee-jerk?  
56  
57  
58  
59  
60

- 1  
2  
3 (2) How can items for a PCK-test be developed on the basis of the identified pupils' errors to  
4  
5 measure teachers' PCK of pupils' errors and ways for dealing with them?  
6  
7  
8 (3) Are these PCK-items reliable and valid for assessing the PCK of pupils' errors?  
9  
10

## 11 12 13 14 15 **Method**

16  
17 The three parts of the study described here are: pupils' errors analysis, development of the PCK-  
18 items as well as the qualitative validation and quantitative analysis of the developed PCK-items.  
19  
20  
21

### 22 23 24 *Participants*

25  
26 An achievement test was given to 461 ninth- and tenth-grade pupils in secondary schools in  
27 Germany—204 male pupils (42.7%) and 249 female pupils (52.1%) (25 pupils without gender  
28 information) with an average age of  $M = 15.14$  years (33 pupils without age information). All the  
29 pupils who took the achievement test (about the reflex arc) had learned about the topic before the  
30 test in their ninth grade.  
31  
32  
33  
34  
35  
36  
37  
38  
39

40  
41 The results of this achievement test were used to develop PCK-items about pupils' errors. To  
42 determine the content validity of these items, five German biology teachers—three teachers are  
43 teaching at the higher secondary school and two at the general-education secondary school—  
44 were interviewed. They had an average age of 43.8 years and an average teaching experience in  
45 biology of 13.40 years. Additionally, all the developed items concerning three different biology  
46 topics and the PCK about pupils' errors were piloted with a larger sample of 65 biology teachers.  
47  
48 They had an average age of 39.54 years and an average teaching experience of 10.50 years (not  
49 including the two years of practicum during the apprenticeship to become a teacher). All the  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 participating teachers—58.5% female and 41.5% male biology teachers from six different federal  
4 states of Germany—answered the PCK-test. Of these participating biology teachers, 72.3%  
5  
6 taught at the higher secondary schools and 27.7% taught at the general-education secondary  
7  
8 schools.  
9  
10

### 11 12 13 14 15 *Achievement Test*

16  
17 In one item of the pupils' achievement test, the pupils had to mark on a drawing of a given  
18 human body, the pathway of the impulse that travels during a *knee-jerk*—the idea was based on  
19  
20 Hammann (2003). The pupils had about five minutes for this item (Figure 1). The expected  
21  
22 answer of the pupils would be the pathway of an *impulse* starting at the *quadriceps muscle*,  
23  
24 travelling to and through the *white* and *gray matter* of the *spinal cord*, and ending in the muscle  
25  
26 again. The pathway from the muscle to the gray matter is the *sensory neuron* and the pathway  
27  
28 back to the muscle is the *motor neuron*, which should be labelled as *afferent* and *efferent nerves*.  
29  
30  
31  
32  
33  
34  
35

36 [Insert fig\_1 about here]  
37  
38  
39  
40

### 41 42 *Development of the PCK-items*

43  
44 Based on the empirically identified and counted error categories of the pupils' answers to the  
45  
46 presented item (see Figure 1), PCK-items for biology teachers were developed in three different  
47  
48 knowledge dimensions: declarative knowledge, procedural knowledge and conditional  
49  
50 knowledge.  
51  
52  
53  
54

55  
56 Figure 2 shows an example of three items—about the biological topic “reflex arc”—as part of the  
57  
58 PCK-test for analysing pupils' errors with respect to the three different knowledge dimensions:  
59  
60

1  
2  
3 Item a) on declarative knowledge; Item b) on conditional knowledge; and Item c) on procedural  
4 knowledge. The most frequently counted pupils' errors category (i.e., category 1 in the results of  
5 this study, see Table 2) was used for the development of this item.  
6  
7  
8  
9

10  
11  
12 [Insert fig\_2 about here]  
13

14  
15  
16  
17 At the beginning of the three items (a, b, c), the teachers got pedagogical background information  
18 (Baxter & Ledermann, 1999) about the pupils: their grade, the point in time of their achievement  
19 test and about how the test looked like. To develop PCK-items which are almost independent of  
20 the CK, the pupils' errors were written explicitly next to the drawing. Because some of the  
21 pupils' errors can be caused by materials given to them in the lesson, this is one aspect asked by  
22 Item 1b). Major aspects of teachers' PCK include not only the knowledge about possible reasons  
23 for pupils' errors but also the knowledge about pupils' upcoming errors and the use of these for  
24 planning future lessons or knowing how someone might react in special error situations (Seifried  
25 & Wuttke, 2010). These aspects are illustrated by the first (Item a) and the third item (Item c), in  
26 Figure 2. The idea of the item development is theoretically based on the empirical result that the  
27 combination of high diagnostic competency, high teachers' judgement accuracy and high  
28 frequency of instructional strategies forms the key for improving instructional quality and  
29 therefore pupils' learning outcomes (Helmke & Schrader, 1987).  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 The coding manual to rate the PCK-items was written on the basis of the literature in science  
51 education, the empirically identified error categories and additional results of schoolbook  
52 analysis. The answers of the teachers to Item a) were coded according to the empirical results of  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the first part of the study by the use of partial credits for exactly correct answers (1), correctly  
4 given frequencies in the interval of  $\pm 10\%$  of the correct answer (0.5) and other answers (0).  
5  
6

7  
8 For coding of answers to Item b), the number of meaningful named reasons was counted. Results  
9  
10 of the document analysis (schoolbooks; school material) as well as literature (Schumacher, 2008;  
11  
12 Seifried & Wuttke, 2010) formed the basis of this coding manual. The idea behind this item was  
13  
14 that the more meaningful reasons the teachers name, the more the teachers might be able to  
15  
16 consider pupils' errors in lesson planning. This coding manual was developed concerning the  
17  
18 qualitative document analysis (schoolbooks; lesson materials; everyday life material like  
19  
20 advertisements). Of course, if a teacher did not name a meaningful reason in the list of examples  
21  
22 of reasons in the coding manual, this was counted with one point as well.  
23  
24  
25

26  
27 The possible reactions of teachers to pupils' errors in the described situation in the biology  
28  
29 lessons (Item c) were coded with one point per each named reaction. Therefore, the listed  
30  
31 possible reactions of the teachers are summarised in four main categories based on theoretical  
32  
33 background: (1) error is used as orientation for further lessons; (2) error is used for other pupils  
34  
35 so that they could learn from others errors' (analysis of errors); (3) error corrected by the pupil  
36  
37 her-/himself so that she/he can learn from her/his errors; and (4) teachers being flexible and  
38  
39 describe the misunderstood topic another way around during the lesson (Jüttner & Neuhaus,  
40  
41 2010; Caspary, 2008; Oser & Spsychiger, 2005; Taber, 2009).  
42  
43  
44  
45

46  
47  
48 All the items concerning "neurobiology" and all PCK components (pupils' errors, models,  
49  
50 experiments) were validated in an additional study. Similar items concerning the other biology  
51  
52 topics (plants and vertebrates) were piloted. For the component "pupils' errors" there are always  
53  
54 three PCK-items per topic according to the knowledge dimensions. Additionally, for the PCK  
55  
56 component "model", always one item per topic—about the reasons of pupils' errors related to the  
57  
58  
59  
60

1  
2  
3 model and its use—could be added to the PCK-items about pupils' errors. So all in all, 12 PCK-  
4  
5 items concerning the knowledge about pupils' errors were created.  
6  
7  
8  
9

### 10 *Data Analysis Methods*

11  
12 The categories of the pupils' errors were identified using the inductive qualitative analysis of  
13 content for 10% of the pupils' answers (Hsieh & Shannon, 2005; Mayring, 2000). The  
14 inductively generated categories were described in the coding manual with which multiple  
15 counting was possible. After the coding of pupils' answers to the shown item by two independent  
16 raters, the relative frequency for each category was computed by using the statistical software  
17 SPSS 19.0.0 (2011). To analyse the interrater reliability for 10% of the sample ( $n = 46$ ) the AC1  
18 statistic (Gwet, 2002; 2008) was calculated in addition to the commonly used Cohen's kappa  
19 value (Landis & Koch, 1977; Shoukri, 2004). This was caused by the paradoxical kappa values  
20 because of the seldom coded categories (Gwet, 2002). Additionally, ten typical German  
21 schoolbooks were analysed to identify information about possible reasons for the categorised  
22 pupils' errors.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 The content validity of the developed PCK-items had been addressed throughout the process of  
42 the development—the empirical data about pupils' errors were used for the creation of the items  
43 of the same biological topic. On the other hand, congruent think-aloud interviews were used  
44 which implicates that the teachers had to communicate aloud their thoughts and simultaneously  
45 write them down on the paper-and-pencil test (Ericsson & Simon, 1993; Leighton, 2009). After  
46 the verbatim transcription of the interviews, the two independent raters coded the interview  
47 transcripts by using a developed coding manual based on the theoretical model (Tepner et al.,  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 sub.) underlying the items. Therefore, the intraclass correlation coefficient (ICC) of the two  
4  
5 independent raters was calculated (Field, 2009; Shrout & Fleiss, 1970).  
6  
7  
8  
9

10 The reliability of the PCK-test was analysed by using the Cronbach's alpha coefficient (Field,  
11  
12 2009). The objectivity was tested by using the intraclass correlation coefficient (ICC) concerning  
13  
14 the coding of teachers' answers to the open-ended PCK-test items (Field, 2009).  
15  
16  
17  
18  
19  
20  
21

## 22 **Results**

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

24  
25 Seven main categories of pupils' errors were identified and counted in the pupils' ( $n = 461$ )  
26  
27 answers. Table 2 shows an overview of the frequencies of the seven categories. Some of the most  
28  
29 often identified error categories are discussed with the following three examples of pupils'  
30  
31 answers (see Figures 3, 4 and 5).  
32  
33  
34  
35  
36  
37  
38

39 [Insert Table\_2 about here]

40  
41  
42  
43 [Insert fig\_3 about here]  
44  
45  
46  
47

48 Figure 3 shows an example of a pupils' answer for four different categories of errors. The most  
49  
50 frequent category 1—*the drawn impulse reaches the brain* (63%)—was identified in this  
51  
52 example. Additionally, two aspects of the second category were identified: *beginning and/or*  
53  
54 *ending of the nerve fibres are not located at the muscle, but at bones* (42%); and *beginning*  
55  
56 *and/or ending of the nerve fibres are not located at the muscle, but in the heel or at the tibia*  
57  
58  
59  
60

1  
2  
3 (14%). Also this example could be used to illustrate the fourth category: *the clearing up centre of*  
4 *the nerves is located in the brain* (26%). The final category identified in this example was  
5  
6 category 5: *nerves are located in the body (stomach) and not in the spinal cord* (17%). In the  
7  
8 drawing, a dot was used by the pupil to represent the change-over point of the pathway in the  
9  
10 brain.  
11  
12  
13

14  
15  
16  
17  
18 [Insert fig\_4 about here]  
19

20  
21  
22 In contrast to the example in Figure 3, the most frequently identified category 1 was not found in  
23  
24 the pupil's answer in Figure 4. In the drawing, the pupil did not connect the brain to the reflex arc  
25  
26 of the knee-jerk. Here, category 2b was identified: *beginning and/or ending of the nerve fibres*  
27  
28 *are not located at the muscle, but at bones* (42%). Moreover, there was *no direct connection*  
29  
30 *between the two nerve fibres so that the reflex arc is not closed*, that is an example for category 6  
31  
32 (15%).  
33  
34  
35

36  
37  
38  
39 [Insert fig\_5 about here]  
40

41  
42  
43 To show examples for every category of pupils' errors, here is the final sample of another pupil's  
44  
45 answer (see Figure 5), where the remaining categories of errors were identified. First of all, the  
46  
47 most frequent category 1 was identified. Furthermore, *the starting- and ending-point of the nerve*  
48  
49 *fibres is in the front of the patella* (category 2a; 47%) as well as category 3: *only one way for the*  
50  
51 *impulse illustrated by one single nerve* (34%) were identified. At last, the least frequent category  
52  
53 7 could be identified: *nerves are not named according to their function, but to their absolute*  
54  
55 *location (brain = afferent nerve; muscle = efferent nerve)* (7%).  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6 For the interrater reliability of this coding, the Cohen's kappa ( $\kappa$ ) for all the categories was .67  
7  
8 ( $p \leq .001$ ,  $n = 462$  categorisations). Because of the distribution of the data, the kappa coefficients  
9  
10 for five categories of errors were very low (e.g., for category 2a,  $\kappa = .35$ ), negative or even not  
11  
12 computable; therefore, the AC1 statistic is described additionally for each category in Table 3  
13  
14 (Gwet, 2008; Schori, Kersten, & Abderhalden, 2006).  
15  
16  
17  
18  
19

20 [Insert Table\_3 about here]  
21  
22  
23

24 The AC1 statistic (Gwet, 2008) showed values from moderate (AC1 (2a) = .50) to almost perfect  
25  
26 (e.g., AC1 (3) = 1) interrater agreement between the two independent raters concerning all the  
27  
28 seven pupils' error categories (see Table 3).  
29  
30  
31  
32  
33

#### 34 *Validation of the Developed PCK-items*

35  
36 The results of the analysis of the coding of the teachers' think-aloud interview transcripts could  
37  
38 demonstrate how often the teachers thought about or used a knowledge dimension to answer the  
39  
40 given PCK-items (see Table 4).  
41  
42  
43  
44  
45

46 [Insert Table\_4 about here]  
47  
48  
49

50 For example, to answer the third item dealing with conditional knowledge on pupils' errors  
51  
52 (PEcond), the five teachers used, on the average, in 90.1% of their thoughts the conditional  
53  
54 knowledge on pupils' errors. The intention of this item was to get information about teachers'  
55  
56 conditional knowledge about pupils' errors. Additionally, while answering this item, they needed  
57  
58  
59  
60

1  
2  
3 or used, on the average, 33% of the declarative knowledge about pupils' errors as well as 16.7%  
4  
5 of declarative content knowledge.  
6  
7

8  
9  
10 The interrater reliability of the two independent raters for the coding of the think-aloud interview  
11 transcripts was found to be significantly high ( $ICC_{\text{unjust}} = .99$ ;  $F_{439,439} = 143.15$ ;  $p < .001$ ).  
12  
13

#### 14 15 16 17 *More Quality Criteria of the PCK-items.*

18  
19 For the internal consistency of the PCK-test scale “declarative knowledge” and the PCK  
20 component “pupils' errors”, a very low Cronbach's alpha value ( $\alpha = .06$ ; 3 items;  $n = 65$ ) was  
21  
22 found. For the two other knowledge dimensions, the Cronbach's alpha values was  $\alpha = .60$  and  
23  
24  
25  
26  
27  $\alpha = .67$  ( $n = 65$ ) (see Table 5).  
28  
29

30  
31  
32 [Insert Table\_5 about here]  
33  
34  
35

36 Finally, the intraclass correlation coefficient (ICC) of the two independent raters for the open-  
37 ended PCK-items was found to be statistically significant ( $ICC_{\text{(unjust)}} = .80$ ;  $F_{588,588} = 5.10$ ;  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
 $p < .001$ ).

### 61 62 63 64 65 66 67 68 69 70 **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

1  
2  
3 and tenth grade German pupils reported in this article might be summed up as three main  
4  
5 categories which might be transferred to other biological topics as well: *wrong locations of the*  
6  
7 *structure; errors concerning the function* and *wrongly used terms and definitions*. The results of  
8  
9 this study showed that most of the analysed pupils' error categories concerning the knee-jerk  
10  
11 were usable for the categorisation of pupils' answers in this topic and that the interrater  
12  
13 agreement (AC1) of the categories ranged from moderate to almost perfect (Gwet, 2002; Landis  
14  
15 & Koch, 1977).  
16  
17  
18  
19

20  
21 All in all, this study was able to show one way of developing PCK-items based on categorisation  
22  
23 of pupils' errors which might enable researchers to develop their topic-specific, practical-  
24  
25 orientated and empirically based PCK-items based on this idea. The quality of the developed  
26  
27 PCK-items was validated by qualitative think-aloud interviews and the quantitative pilot phase.  
28  
29 The results of these analyses demonstrated the development of a valid, reliable and objective  
30  
31 PCK-test instrument for testing teachers' PCK concerning pupils' errors in biology.  
32  
33  
34  
35

36 The developed items in the PCK-test have curricular content validity because they were  
37  
38 developed after curricular reviews and many meetings with experienced teachers. Additionally,  
39  
40 the empirical data of the pupils' errors gave more content validity. Finally, the think-aloud  
41  
42 interviews demonstrated that teachers mostly thought of or used the kind of knowledge  
43  
44 (declarative, procedural or conditional) for answering the item which was intended to be  
45  
46 measured by the respective item. The ICC of the interview transcript coding showed a  
47  
48 satisfactory interrater agreement between the two independent raters (Field, 2009). In the end of  
49  
50 all the interviews, four follow-up questions about the items were also asked. One of these  
51  
52 questions asked the teacher if their answers to the items might represent their personal way of  
53  
54 enactment in classroom situations or their way of thinking concerning example models in biology  
55  
56  
57  
58  
59  
60

1  
2  
3 lessons. This could be seen as face validity, because all the interviewed teachers said something  
4 similar to what Teacher K said: “My answers represent what I strive to do”.

5  
6  
7  
8 The internal consistency of the PCK-test (its Cronbach’s alpha values) showed satisfactory  
9 results except for the declarative knowledge of pupils’ errors—about teachers’ prediction of how  
10 often the given pupils’ error might arise after the learning process. One reason might be that  
11 teachers often wrongly predict their pupils’ learning outcomes and because they have over-  
12 expectations of their pupils’ learning outcomes (Helmke & Schrader, 1987; Hosenfeld, Helmke  
13 & Schrader, 2002).

14  
15  
16  
17 The objectivity of the PCK-items was also confirmed with an acceptable intraclass correlation  
18 coefficient (ICC).

19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29 Due to the small sample size of this study for more detailed statistical analysis (e.g., factor  
30 analysis), further studies with samples of about 200 teachers will be conducted to test if all the  
31 developed items conform to the three dimensional theoretical model of the current project  
32 *ProwiN* (Tepner et al., sub.). Additionally, the items were translated into English and will be  
33 given to American teachers in our further studies. The results of the future think-aloud interviews  
34 might show the possible use and transferability of this study’s development ideas to other  
35 countries and cultures.

36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48 Teachers’ knowledge about pupils’ possible errors—which may arise in a lesson or about  
49 possible reasons for these errors—is important for teachers because this knowledge has different  
50 advantages for teachers. On the one hand, it can be used as an orientation during the planning  
51 process of lessons (Schumacher, 2008; Seifried & Wuttke, 2010); and meaningful arrangement of  
52 materials, figures and models can also result. On the other hand, the information can be used for  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the training of pupils' error analysis; pupils have the possibility to learn the so-called *negative*  
4 *knowledge* (Oser, Hascher, & Spychiger, 1999). In the analysis of errors made by anonymous  
5  
6  
7  
8 pupils, the pupils have to change their point of view to that of a corrector. This change is the best  
9  
10 opportunity for learning from errors (Schumacher, 2008). Therefore, empirically analysed and  
11  
12 categorised pupils' errors, as reported in this article, are important for teachers. In addition, these  
13  
14 results can be used for test items construction in science education research as well as in teacher  
15  
16  
17  
18 education.

19  
20  
21  
22 To conclude, the method of developing the PCK-items demonstrated in our study was able to test  
23  
24 teachers' knowledge by its connectedness to praxis, as well as to relate pupils' performance in  
25  
26 achievement tests directly to teachers' knowledge about their pupils' errors. So it will be possible  
27  
28 to construct tests for pupils and teachers for analysing the correlation between pupils' knowledge  
29  
30 and teachers' PCK concerning pupils' understanding.  
31  
32  
33  
34  
35

36  
37 As for research, the aim will be to analyse the PCK of teachers and to see if their knowledge  
38  
39 correlates with how they prepare and teach the subject matter in different biological topics as well  
40  
41 as with their job experience. So far, mathematical pre-service teachers' experience has been  
42  
43 analysed in longitudinal studies (e.g. Tatto et al., 2008). Additionally, there exist a few qualitative  
44  
45 studies with a smaller sample size evaluating the development of teachers' professional  
46  
47 knowledge (Henze & van Driel, 2009; Friedrichsen, Abell, Pareja, Brown, Lankford,  
48  
49 &Volkmann, 2007). In research there are different results concerning the influence of teachers'  
50  
51 experience. Friedrichsen et al. (2007) found that the development of science teachers'  
52  
53 pedagogical knowledge was supported by experience but the PCK development was not. Further  
54  
55  
56  
57  
58 research concerning the analysis and conceptualisation of general pedagogical knowledge is  
59  
60

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 activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133-180.

- 1  
2  
3 Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content  
4 knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content*  
5 *knowledge* (pp. 147-161). Dordrecht: Kluwer.  
6  
7  
8  
9  
10 Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des  
11  
12 erziehungswissenschaftlichen Wissens angehender Lehrkräfte [Assessment of student  
13 teachers' pedagogical knowledge]. In S. Blömeke, G. Kaiser & R. Lehmann (Eds.),  
14 *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen*  
15 *und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste*  
16 *Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 171–194). Münster, Germany:  
17 Waxmann.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27 Bouvier, A.(1987). The right to make mistakes. *For the Learning of Mathematics*, 7(3), 17–25.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 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.

- 1  
2  
3 de Laughter, J. E., & Stein, S. (1998). Preconceptions about earth science among pupils in an  
4  
5 introductory course. *Eos*, 79, 429-434.  
6  
7  
8 Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory  
9  
10 and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual*  
11  
12 *change* (pp. 629-646). New York, London: Routledge.  
13  
14  
15 Driver, R.P. (1981). Pupils' alternative frameworks in science. *European Journal of Science*  
16  
17 *Education*, 3(1), 93-101.  
18  
19  
20 Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: The MIT Press.  
21  
22  
23 Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.  
24  
25 Fischer, H.E., Borowski, A., & Tepner, O. (in press). Professional knowledge of science teachers;  
26  
27 In B. Fraser; K. Tobin, & C. McRobbie (Eds.). *Second International Handbook of Science*  
28  
29 *Education*, New York: Springer.  
30  
31  
32 Friedrichsen, P., Abell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does  
33  
34 teaching experience matter? Examining biology teachers' prior knowledge for teaching in an  
35  
36 alternative certification program. *Journal of Research in Science Teaching*, 46, 357-383.  
37  
38  
39 Gilbert, J., Osborne, R., & Fensham, P. (1982). Children's science and its consequences for  
40  
41 teaching. *Science Education*, 66, 623-33.  
42  
43  
44 Gwet, K. L. (2008). *Interrater Reliability*. Wiley Encyclopedia of Clinical Trials, Copyright 2008  
45  
46 John Wiley & Sons, Inc.  
47  
48 Gwet, K. L. (2002). Kappa Statistic is not satisfactory for assessing the extent of agreement  
49  
50 between raters. *Statistical Methods For Inter-Rater Reliability Assessment*, 1, 1-5. Retrieved  
51  
52 from [http://agreestat.com/research\\_papers/kappa\\_statistic\\_is\\_not\\_satisfactory.pdf](http://agreestat.com/research_papers/kappa_statistic_is_not_satisfactory.pdf) [Date  
53  
54 accessed: 25 May. 2011].  
55  
56  
57  
58  
59  
60 Hammann, M. (2003). Aus Fehlern lernen. [Learning by errors], *Unterricht Biologie*, 287, 31-35.



- 1  
2  
3 Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of PCK. *Teachers*  
4 *and Teaching: Theory and Practice*, 11, 273–292.  
5  
6  
7  
8 Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher  
9 judgment accuracy on achievement. *Teaching and Teacher Education*, 3, 91-98  
10  
11  
12 Helm, H. (1980). Misconceptions in physics amongst South African pupils. *Physics Education*,  
13 15, 92-7.  
14  
15  
16  
17 Henze, I., Driel van, J. H., & Verloop, N. (2008). Development of experienced science teachers'  
18 pedagogical content knowledge of models of the solar system and the universe. *International*  
19 *Journal of Science Education*, 30(10), 1321-1342.  
20  
21  
22  
23  
24 Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's  
25 mathematics professional development institutes. *Journal for Research in Mathematics*  
26 *Education*, 35 (5), 330-351.  
27  
28  
29  
30  
31 Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical  
32 knowledge for teaching on student achievement. *American Educational Research Journal*,  
33 42, 371-406.  
34  
35  
36  
37  
38 Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge:  
39 Conceptualizing and measuring teachers' topic-specific knowledge of pupils. *Journal for*  
40 *Reserach in Mathematics Education*, 39, 372-400.  
41  
42  
43  
44  
45 Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts-  
46 und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Lehr-  
47 Lern-Studie SALVE. [Diagnotic competency: teaching and learning relevant  
48 to student characteristics and the teachers' prediction in the teaching-learning study SALVE],  
49 In M. Prenzel & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische*  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 *Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen.*  
4  
5 *Zeitschrift für Pädagogik*, 45. Beiheft (pp. 65-82). Beltz: Weinheim.  
6  
7

8 Houssart, J., & Weller, B. (1999). Identifying and dealing with misconceptions and errors in  
9  
10 primary mathematics: Student teachers record their experiences. *Mathematics Education*  
11  
12 *Review*, 11, 46-58.  
13

14  
15 Hsieh, H.-F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis.  
16  
17 *Qualitative Health Research*, 15(9), 1277-1288.  
18

19  
20 Jüttner, M., & Neuhaus, B. (2011). Vom Schülerfehler zum PCK – Test. Ansätze zur  
21  
22 Entwicklung von PCK – Items. [A PCK-test based on students' errors. Rudiments for  
23  
24 developing PCK items.] In U. Harms & I. Mackensen-Friedrichs (Eds.), *Lehr- und*  
25  
26 *Lernforschung in der Biologiedidaktik. Band 4* (pp. 27-39). Studien Verlag: Innsbruck.  
27  
28

29  
30 Jüttner, M. & Neuhaus, B. (2010). Using empirically analyzed pupils' errors to develop a PCK-  
31  
32 Test. In M.F. Taşar & G. Çakmakçı (Eds.), *Contemporary science education research:*  
33  
34 *preservice and inservice teacher education* (pp. 331-340). Ankara, Turkey: Pegem Akademi.  
35

36  
37 Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential  
38  
39 for progress. *Studies in Science Education*, 45(29), 169-204.  
40

41  
42 Krüger, D. (2007). Die Conceptuale Change-Theorie. [The conceptual change theory]. In D.  
43  
44 Krüger & H. Vogt (Eds.), *Handbuch der Theorien in der biologiedidaktischen Forschung*  
45  
46 (pp. 81-92). Berlin: Springer Verlag.  
47

48  
49 Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical  
50  
51 data. *Biometrics*, 33 (1), 159–174.  
52

53  
54 Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of  
55  
56 pedagogical content knowledge. *International Journal of Science Education*, 30, 1343-1363.  
57  
58  
59  
60

- 1  
2  
3 Leighton, J. P.(2009). Two Types of Think Aloud Interviews for Educational Measurement:  
4  
5 Protocol and Verbal Analysis. *Paper presented at the National Council on Measurement in*  
6  
7  
8 *Education (NCME), CA.*  
9
- 10 Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge  
11  
12 zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. [It depends on  
13  
14 the teacher. Empirical evidences of the connection between teachers' competence, teachers'  
15  
16 acting and the learning of the pupils], *Zeitschrift für Pädagogik*, 52 (51), 47-70.  
17  
18
- 19 Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in  
20  
21 science teacher education. *International Journal of Science Education*, 30 (10), 1301-1320.  
22  
23
- 24 Magnusson, S., & Krajcik, J. & Borke, H. (1999). Nature, sources and development of  
25  
26 pedagogical content knowledge for science teaching. In: Gess-Newsome & Lederman (Eds.),  
27  
28 *Examining pedagogical content knowledge: The construct and its implications for science*  
29  
30 *education* (pp. 95-131). Dordrecht. Kluwer Academic Publishers.  
31  
32
- 33 Mayring, P. (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung / Forum:  
34  
35 Qualitative Social Research, North America, 1. Available at: [http://www.qualitative-](http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385)  
36  
37 [research.net/index.php/fqs/article/view/1089/2385](http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385). [Date accessed: 26 Oct. 2010].  
38  
39
- 40 Nebah, S. J. (2002). *Do people learn from mistakes? A pedagogical study and analysis of*  
41  
42 *autobiographies and interviews*. Dissertation am Departement Erziehungswissenschaften der  
43  
44 Universität Fribourg.  
45  
46
- 47 Novak, G.S. (1977). *Representations of knowledge in a program for solving physics problems.*  
48  
49 *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.*  
50  
51  
52  
53 Cambridge, MA: MIT Press.  
54
- 55 Oser, F., Hascher, T., & Spsychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie der  
56  
57 „negativen“ Wissens. [Learning out of errors. About the psychology of the "negative"  
58  
59  
60

1  
2  
3 knowledge]. In W. Althof (Ed.), *Fehlerwelten: vom Fehlermachen und Lernen aus Fehlern*  
4  
5 (pp. 11-43). Opladen: Leske + Budrich.

6  
7  
8 Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens*  
9  
10 *und zur Praxis der Fehlerkultur.* [Learning is painful. About the theory of the negative  
11  
12 knowledge and about the practice of the culture of errors]. Weinheim: Beltz.

13  
14  
15 Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary*  
16  
17 *Educational Psychology*, 8, 293-316.

18  
19  
20 Park, S., Jang, J.-Y., & Chen, Y.-C. (2010). Is Pedagogical Content Knowledge (PCK) necessary  
21  
22 for reformed science teaching?: Evidence from an empirical study. *Research in Science*  
23  
24 *Education*, 41, 245–260. doi: 10.1007/s11165-009-9163-8

25  
26  
27 Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of Pedagogical Content  
28  
29 Knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals.  
30  
31 *Research in Science Education*, 38, 261-284.

32  
33  
34 Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Measuring declarative and  
35  
36 reflective components of biology teachers' pedagogical content knowledge. In M. F. Tasar &  
37  
38 G. Cakmakci (Eds.), *Contemporary science education research: Teaching* (pp. 71-77).  
39  
40 Ankara, Turkey: Pegem Akademi.

41  
42  
43 Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L. et al. (2007). *The*  
44  
45 *Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries.*  
46  
47 *Mathematics Teaching in the 21st Century (MT21)* (Rep. No. MSU Center for Research in  
48  
49 Mathematics and Science Education).

50  
51  
52  
53 Schori, E., Kersten, B. & Abderhalden, C. (2006). Wie reliabel ist der Fragebogen zur Analyse  
54  
55 der Pflegedokumentation IzEP-Dok© im Instrument zur Erfassung von Pflegesystemen  
56  
57 IzEP©? [How Reliable is the Questionnaire for the Analysis of Nursing Documentation  
58  
59  
60

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 Fehlerkultur. [An open play. Learning by errors and the development of the culture of errors]. In R.

1  
2  
3 Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp.  
4  
5 25-48). Freiburg: Herder.

6  
7  
8 Taber, K. S. (2009). *Progressing science education: Constructing the scientific research*  
9  
10 *programme into the contingent nature of learning science*. Dordrecht: Springer.

11  
12  
13 Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: from a scholarly  
14  
15 approach to teaching to the scholarship of teaching. *Cell Biology Education*, 4, 1-6.

16  
17  
18 Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher*  
19  
20 *Education and Development Study in Mathematics (TEDS-M): Conceptual framework*. East  
21  
22 Lansing, MI: Teacher Education and Development International Study Center, College of  
23  
24 Education, Michigan State University.

25  
26  
27 Tepner, O., Borowski, A., Fischer, H.E., Jüttner, M., Kirschner, S., Leutner, D. et al. (submitted).  
28  
29 Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von  
30  
31 Lehrkräften in den Naturwissenschaften. [Theoretical model for the development of test-  
32  
33 items to collect data about professional knowledge of science teachers], *Zeitschrift der*  
34  
35 *Naturwissenschaften*.

36  
37  
38 Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students'  
39  
40 misconceptions in science. *International Journal of Science Education*, 10 (2), 159-169.

41  
42  
43 Tsai, C.-C., & Chou, C. (2002). Diagnosing pupils' alternative conceptions in science. *Journal of*  
44  
45 *Computer Assisted Learning*, 18, 157-165.

46  
47  
48 Tsamir, P., Rasslan, S., & Dreyfus, T. (2006). Prospective teachers' reactions to Right-or-Wrong  
49  
50 tasks: The case of derivatives of absolute value functions. *Journal of Mathematical*  
51  
52 *Behavior*, 25, 240-251.

53  
54  
55 van Driel, J. H., Verloop, N., & Vos, W. d. (1998). Developing science teachers' Pedagogical  
56  
57 Content Knowledge. *Journal of Research in Science Teaching*, 35, 673-695.

1  
2  
3 van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in  
4  
5 science education: The role of teachers' practical knowledge. *Journal of Research in Science*  
6  
7  
8 *Teaching*, 38, 137-158.  
9

10  
11  
12  
13  
14  
15 \* the presented project is funded by the Federal Ministry of Education and Research; it is a cooperation project embedded in the  
16  
17 framework program of 'empirical research in education' (01JH0904).  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# 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.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For Peer Review Only

## 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 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

1  
2  
3 pedagogical knowledge (PK). Over the past decade, many research projects have used these three  
4  
5 main categories of teachers' professional knowledge to provide a theoretical background. Unlike  
6  
7 most of the current studies on mathematics teachers (e.g., Abell, 2007; Baumert et al., 2010; Hill  
8  
9 & Ball, 2004; Lipowsky, 2006; Park & Oliver, 2008; Schmidt et al., 2007), the project *ProwiN*  
10  
11 (the German acronym for *Professional Knowledge in Science*) (Fischer, Borowski, & Tepner, in  
12  
13 press) is a first step to close the gap in research about teachers' professional knowledge in science  
14  
15 education (Abell, 2007). Three cooperating German Universities analysed the PCK, the CK and  
16  
17 the PK of science teachers based on a theoretical model (Tepner et al., submitted) and the results  
18  
19 of researchers-developed test instruments. The study presented here is part of the project *ProwiN*  
20  
21 with a focus on one category of the professional knowledge—the PCK of biology teachers.  
22  
23  
24  
25  
26  
27  
28  
29

### 30 *Theoretical Conceptualisations of the Pedagogical Content Knowledge (PCK)*

31  
32  
33 Most studies on PCK focus on one of the following PCK components: “knowledge about pupils’  
34  
35 understanding”, “curricular knowledge”, “knowledge about instructional strategies and  
36  
37 representation” (cf. Kind, 2009; Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008; van  
38  
39 Driel, Verloop, & de Vos, 1998). Illustrated here is the development of an instrument for  
40  
41 measuring biology teachers' PCK with a special focus on one PCK component—the pupils’  
42  
43 errors and their understanding of biology. Of special interest to this study is the *MT21*-study  
44  
45 (Schmidt et al., 2007) which analysed two sub-competencies of mathematic teachers' diagnostic  
46  
47 competency: recognition of pupils' misconceptions and criteria-guided assessment of pupils'  
48  
49 solutions to problems (Schwarz, Wissmach, & Kaiser, 2008).  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 The PCK in this study (part of the project *ProwiN*) was generally defined according to Shulman  
4 (1986) as knowledge of structuring and describing the teaching of the content that would be  
5 necessary for preparing the subject matter so that it is comprehensible for pupils (Fischer,  
6 Borowski, & Tepner, in press). The most often used PCK components in the literature—  
7 knowledge about pupils' understanding, instructional strategies and representations (Park &  
8 Oliver, 2008; Schmelzing et al., 2010)—were chosen for the PCK-model of the project (Tepner et  
9 al., sub.).  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20  
21  
22 The knowledge about pupils' understanding will be highlighted. This PCK component is defined  
23 throughout pupils' errors as well as the knowledge about the thinking of the pupils in solving  
24 problems. In addition to the three PCK components of the project (pupils' errors, models and  
25 experiments), three knowledge dimensions were defined: declarative, procedural and conditional  
26 knowledge (Alexander, Schallert, & Hare, 1991; de Jong & Ferguson-Hessler, 1996; Paris,  
27 Lipson, & Wixson, 1983). These knowledge dimensions form the required cognitive aspect of  
28 teachers' professional knowledge—here the PCK conceptualisation (Hill, Ball, & Schilling,  
29 2008). Moreover three different biology topics that teachers have to know about—neurobiology,  
30 plants and vertebrates—were selected from the curriculum of German schools for the  
31 development of the PCK-items. In this article, we use the human reflex arc to demonstrate the  
32 idea of developing PCK-items concerning “pupils' errors” and the three different knowledge  
33 dimensions (declarative, procedural and conditional).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

#### 54 *Pupils' Errors*

55  
56  
57  
58  
59  
60

1  
2  
3 In science education, there are many definitions of different terms which have been used for  
4 describing pupils' knowledge (Taber, 2009). Sometimes, some terms could be used  
5 synonymously and sometimes one and the same term is discussed in different ways. Table 1  
6 shows an overview of the exemplary literature on some of the accepted terms about pupils'  
7 knowledge in science that were categorised and combined into two different groups for defining  
8 "pupils' errors" in our study.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20 [Insert Table\_1 about here]  
21  
22  
23

24 *Conceptions and prior knowledge.* In the literature about pupils' scientifically incorrect  
25 conceptions, two main aspects have become apparent—pupils can bring their ideas with them to  
26 the classroom from their everyday life or they can develop their ideas during the lesson and/or  
27 because of the topic being taught in the lesson. Accordingly, the terms can be classified into two  
28 groups. (1) *Conceptions*—defined as a way of pupils' thinking which they established on their  
29 own depending on their experience in everyday life but not consistent with teachers' and  
30 scientists' ideas—have often been called *misconceptions*, *preconceptions*, *alternative*  
31 *conceptions*, *alternative frameworks* or *children's science* (Treagust, 1988). (2) *Prior knowledge*  
32 (Allen, 2005)—described as knowledge before pupils learn something about a topic—has been  
33 studied since the 1970s. *Knowledge* is used as a tightly-defined term and *idea*, *belief*, as well as  
34 *concept* are sometimes used as synonyms (Taber, 2009). But knowledge differs from  
35 conceptions, because knowledge of pupils can be described as a set of conceptions.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

52 *Pupils' errors.* Concurrent to studies on "conceptions or misconceptions", research in  
53 pedagogical psychology explores "*pupils' errors*" (Oser, Hascher, & Spychiger, 1999; Seifried &  
54 Wuttke, 2010; Spychiger, 2008). Pupils' errors are the out coming signals of their difficulties or  
55  
56  
57  
58  
59  
60

1  
2  
3 misunderstandings in lessons. The “errors” are different from the “misconceptions”, because  
4  
5 misconceptions could be defined as scientifically incorrect knowledge about something which  
6  
7 can and should influence pupils’ conceptual learning in the lesson (Tanner & Allen, 2005). The  
8  
9 meaning between neutral and non-neutral terms can be distinguished, for example, “error” or  
10  
11 “knowledge” (Taber, 2009), which indicates the possibility of saying if something is wrong or  
12  
13 right according to the current scientific research on the topic. But there are problems within the  
14  
15 definition as well—what is wrong and what is right in science (Cokelez, Dumon, & Taber,  
16  
17 2008)? Using these non-neutral—often negatively understood—terms, researchers have to define  
18  
19 them clearly for the terms not being misunderstood (Oser, Hascher, & Spychiger, 1999). On the  
20  
21 contrary, a neutral term will be, for example, “children’s ideas” and “alternative conceptions”  
22  
23 (Taber, 2009), which are defined more openly to include most ideas of pupils.  
24  
25  
26  
27  
28  
29  
30  
31

32 In this study, the term “errors” was used because we were interested in errors of the ninth and  
33  
34 tenth grade pupils after their learning process in biology in the ninth grade. The focus was on  
35  
36 errors of pupils because these errors could be seen as an indicator of the effectiveness of  
37  
38 instruction (Seidel & Shavelson, 2007). Pupils’ errors are an important component of teachers’  
39  
40 PCK on how to handle errors and to examine what might be the reasons for such errors (Baumert  
41  
42 et al., 2010; Hill, Rowan, & Loewenberg Ball, 2005; Lipowsky, 2006; Seifried, & Wuttke, 2010).  
43  
44  
45  
46  
47

#### 48 *Aims*

49  
50 Based on the theoretical background, the following were the research questions and aims of our  
51  
52 study:  
53  
54

- 55 (1) What are the typical pupils’ errors about the reflex arc of the knee-jerk?  
56  
57  
58  
59  
60

- 1  
2  
3 (2) How can items for a PCK-test be developed on the basis of the identified pupils' errors to  
4  
5 measure teachers' PCK of pupils' errors and ways for dealing with them?  
6  
7  
8 (3) Are these PCK-items reliable and valid for assessing the PCK of pupils' errors?  
9

## 10 11 12 13 14 15 **Method**

16  
17 The three parts of the study described here are: pupils' errors analysis, development of the PCK-  
18 items as well as the qualitative validation and quantitative analysis of the developed PCK-items.  
19  
20  
21

### 22 23 24 *Participants*

25  
26 An achievement test was given to 461 ninth- and tenth-grade pupils in secondary schools in  
27 Germany—204 male pupils (42.7%) and 249 female pupils (52.1%) (25 pupils without gender  
28 information) with an average age of  $M = 15.14$  years (33 pupils without age information). All the  
29 pupils who took the achievement test (about the reflex arc) had learned about the topic before the  
30 test in their ninth grade.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 The results of this achievement test were used to develop PCK-items about pupils' errors. To  
42 determine the content validity of these items, five German biology teachers—three teachers are  
43 teaching at the higher secondary school and two at the general-education secondary school—  
44 were interviewed. They had an average age of 43.8 years and an average teaching experience in  
45 biology of 13.40 years. Additionally, all the developed items concerning three different biology  
46 topics and the PCK about pupils' errors were piloted with a larger sample of 65 biology teachers.  
47  
48 They had an average age of 39.54 years and an average teaching experience of 10.50 years (not  
49 including the two years of practicum during the apprenticeship to become a teacher). All the  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 participating teachers—58.5% female and 41.5% male biology teachers from six different federal  
4 states of Germany—answered the PCK-test. Of these participating biology teachers, 72.3%  
5  
6 taught at the higher secondary schools and 27.7% taught at the general-education secondary  
7  
8 schools.  
9  
10

### 11 12 13 *Achievement Test*

14  
15  
16  
17 In one item of the pupils' achievement test, the pupils had to mark on a drawing of a given  
18 human body, the pathway of the impulse that travels during a *knee-jerk*—the idea was based on  
19  
20 Hammann (2003). The pupils had about five minutes for this item (Figure 1). The expected  
21  
22 answer of the pupils would be the pathway of an *impulse* starting at the *quadriceps muscle*,  
23  
24 travelling to and through the *white* and *gray matter* of the *spinal cord*, and ending in the muscle  
25  
26 again. The pathway from the muscle to the gray matter is the *sensory neuron* and the pathway  
27  
28 back to the muscle is the *motor neuron*, which should be labelled as *afferent* and *efferent nerves*.  
29  
30  
31  
32  
33  
34  
35

36 [Insert fig\_1 about here]  
37  
38  
39  
40

### 41 *Development of the PCK-items*

42  
43 Based on the empirically identified and counted error categories of the pupils' answers to the  
44 presented item (see Figure 1), PCK-items for biology teachers were developed in three different  
45  
46 knowledge dimensions: declarative knowledge, procedural knowledge and conditional  
47  
48 knowledge.  
49  
50  
51  
52  
53  
54

55 Figure 2 shows an example of three items—about the biological topic “reflex arc”—as part of the  
56  
57 PCK-test for analysing pupils' errors with respect to the three different knowledge dimensions:  
58  
59  
60



1  
2  
3 Item a) on declarative knowledge; Item b) on conditional knowledge; and Item c) on procedural  
4  
5 knowledge. The most frequently counted pupils' errors category (i.e., category 1 in the results of  
6  
7 this study, see Table 2) was used for the development of this item.  
8  
9

10  
11  
12 [Insert fig\_2 about here]  
13

14  
15  
16  
17 At the beginning of the three items (a, b, c), the teachers got pedagogical background information  
18  
19 (Baxter & Ledermann, 1999) about the pupils: their grade, the point in time of their achievement  
20  
21 test and about how the test looked like. To develop PCK-items which are almost independent of  
22  
23 the CK, the pupils' errors were written explicitly next to the drawing. Because some of the  
24  
25 pupils' errors can be caused by materials given to them in the lesson, this is one aspect asked by  
26  
27 Item 1b). Major aspects of teachers' PCK include not only the knowledge about possible reasons  
28  
29 for pupils' errors but also the knowledge about pupils' upcoming errors and the use of these for  
30  
31 planning future lessons or knowing how someone might react in special error situations (Seifried  
32  
33 & Wuttke, 2010). These aspects are illustrated by the first (Item a) and the third item (Item c), in  
34  
35 Figure 2. The idea of the item development is theoretically based on the empirical result that the  
36  
37 combination of high diagnostic competency, high teachers' judgement accuracy and high  
38  
39 frequency of instructional strategies forms the key for improving instructional quality and  
40  
41 therefore pupils' learning outcomes (Helmke & Schrader, 1987).  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 The coding manual to rate the PCK-items was written on the basis of the literature in science  
52  
53 education, the empirically identified error categories and additional results of schoolbook  
54  
55 analysis. The answers of the teachers to Item a) were coded according to the empirical results of  
56  
57  
58  
59  
60

1  
2  
3 the first part of the study by the use of partial credits for exactly correct answers (1), correctly  
4 given frequencies in the interval of  $\pm 10\%$  of the correct answer (0.5) and other answers (0).

5  
6  
7  
8 For coding of answers to Item b), the number of meaningful named reasons was counted. Results  
9  
10 of the document analysis (schoolbooks; school material) as well as literature (Schumacher, 2008;  
11 Seifried & Wuttke, 2010) formed the basis of this coding manual. The idea behind this item was  
12 that the more meaningful reasons the teachers name, the more the teachers might be able to  
13 consider pupils' errors in lesson planning. This coding manual was developed concerning the  
14 qualitative document analysis (schoolbooks; lesson materials; everyday life material like  
15 advertisements). Of course, if a teacher did not name a meaningful reason in the list of examples  
16 of reasons in the coding manual, this was counted with one point as well.

17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27 The possible reactions of teachers to pupils' errors in the described situation in the biology  
28 lessons (Item c) were coded with one point per each named reaction. Therefore, the listed  
29 possible reactions of the teachers are summarised in four main categories based on theoretical  
30 background: (1) error is used as orientation for further lessons; (2) error is used for other pupils  
31 so that they could learn from others errors' (analysis of errors); (3) error corrected by the pupil  
32 her-/himself so that she/he can learn from her/his errors; and (4) teachers being flexible and  
33 describe the misunderstood topic another way around during the lesson (Jüttner & Neuhaus,  
34 2010; Caspary, 2008; Oser & Spychiger, 2005; Taber, 2009).

35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48 All the items concerning "neurobiology" and all PCK components (pupils' errors, models,  
49 experiments) were validated in an additional study. Similar items concerning the other biology  
50 topics (plants and vertebrates) were piloted. For the component "pupils' errors" there are always  
51 three PCK-items per topic according to the knowledge dimensions. Additionally, for the PCK  
52 component "model", always one item per topic—about the reasons of pupils' errors related to the  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 model and its use—could be added to the PCK-items about pupils' errors. So all in all, 12 PCK-  
4  
5 items concerning the knowledge about pupils' errors were created.  
6  
7  
8  
9

### 10 *Data Analysis Methods*

11  
12 The categories of the pupils' errors were identified using the inductive qualitative analysis of  
13  
14 content for 10% of the pupils' answers (Hsieh & Shannon, 2005; Mayring, 2000). The  
15  
16 inductively generated categories were described in the coding manual with which multiple  
17  
18 counting was possible. After the coding of pupils' answers to the shown item by two independent  
19  
20 raters, the relative frequency for each category was computed by using the statistical software  
21  
22 SPSS 19.0.0 (2011). To analyse the interrater reliability for 10% of the sample ( $n = 46$ ) the AC1  
23  
24 statistic (Gwet, 2002; 2008) was calculated in addition to the commonly used Cohen's kappa  
25  
26 value (Landis & Koch, 1977; Shoukri, 2004). This was caused by the paradoxical kappa values  
27  
28 because of the seldom coded categories (Gwet, 2002). Additionally, ten typical German  
29  
30 schoolbooks were analysed to identify information about possible reasons for the categorised  
31  
32 pupils' errors.  
33  
34  
35  
36  
37  
38  
39  
40

41 The content validity of the developed PCK-items had been addressed throughout the process of  
42  
43 the development—the empirical data about pupils' errors were used for the creation of the items  
44  
45 of the same biological topic. On the other hand, congruent think-aloud interviews were used  
46  
47 which implicates that the teachers had to communicate aloud their thoughts and simultaneously  
48  
49 write them down on the paper-and-pencil test (Ericsson & Simon, 1993; Leighton, 2009). After  
50  
51 the verbatim transcription of the interviews, the two independent raters coded the interview  
52  
53 transcripts by using a developed coding manual based on the theoretical model (Tepner et al.,  
54  
55  
56  
57  
58  
59  
60

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*

1  
2  
3 (14%). Also this example could be used to illustrate the fourth category: *the clearing up centre of*  
4 *the nerves is located in the brain* (26%). The final category identified in this example was  
5  
6  
7  
8 category 5: *nerves are located in the body (stomach) and not in the spinal cord* (17%). In the  
9  
10 drawing, a dot was used by the pupil to represent the change-over point of the pathway in the  
11  
12 brain.  
13

14  
15  
16  
17  
18 [Insert fig\_4 about here]  
19

20  
21  
22 In contrast to the example in Figure 3, the most frequently identified category 1 was not found in  
23  
24 the pupil's answer in Figure 4. In the drawing, the pupil did not connect the brain to the reflex arc  
25  
26 of the knee-jerk. Here, category 2b was identified: *beginning and/or ending of the nerve fibres*  
27  
28 *are not located at the muscle, but at bones* (42%). Moreover, there was *no direct connection*  
29  
30 *between the two nerve fibres so that the reflex arc is not closed*, that is an example for category 6  
31  
32 (15%).  
33  
34  
35

36  
37  
38  
39 [Insert fig\_5 about here]  
40

41  
42  
43 To show examples for every category of pupils' errors, here is the final sample of another pupil's  
44  
45 answer (see Figure 5), where the remaining categories of errors were identified. First of all, the  
46  
47 most frequent category 1 was identified. Furthermore, *the starting- and ending-point of the nerve*  
48  
49 *fibres is in the front of the patella* (category 2a; 47%) as well as category 3: *only one way for the*  
50  
51 *impulse illustrated by one single nerve* (34%) were identified. At last, the least frequent category  
52  
53  
54  
55 7 could be identified: *nerves are not named according to their function, but to their absolute*  
56  
57 *location (brain = afferent nerve; muscle = efferent nerve)* (7%).  
58  
59  
60

1  
2  
3  
4  
5  
6 For the interrater reliability of this coding, the Cohen's kappa ( $\kappa$ ) for all the categories was .67  
7  
8 ( $p \leq .001$ ,  $n = 462$  categorisations). Because of the distribution of the data, the kappa coefficients  
9  
10 for five categories of errors were very low (e.g., for category 2a,  $\kappa = .35$ ), negative or even not  
11  
12 computable; therefore, the AC1 statistic is described additionally for each category in Table 3  
13  
14 (Gwet, 2008; Schori, Kersten, & Abderhalden, 2006).  
15  
16  
17  
18  
19

20 [Insert Table\_3 about here]  
21  
22  
23

24 The AC1 statistic (Gwet, 2008) showed values from moderate (AC1 (2a) = .50) to almost perfect  
25  
26 (e.g., AC1 (3) = 1) interrater agreement between the two independent raters concerning all the  
27  
28 seven pupils' error categories (see Table 3).  
29  
30  
31  
32  
33

#### 34 *Validation of the Developed PCK-items*

35  
36 The results of the analysis of the coding of the teachers' think-aloud interview transcripts could  
37  
38 demonstrate how often the teachers thought about or used a knowledge dimension to answer the  
39  
40 given PCK-items (see Table 4).  
41  
42  
43  
44  
45

46 [Insert Table\_4 about here]  
47  
48  
49

50 For example, to answer the third item dealing with conditional knowledge on pupils' errors  
51  
52 (PEcond), the five teachers used, on the average, in 90.1% of their thoughts the conditional  
53  
54 knowledge on pupils' errors. The intention of this item was to get information about teachers'  
55  
56 conditional knowledge about pupils' errors. Additionally, while answering this item, they needed  
57  
58  
59  
60

1  
2  
3 or used, on the average, 33% of the declarative knowledge about pupils' errors as well as 16.7%  
4  
5 of declarative content knowledge.  
6  
7

8  
9  
10 The interrater reliability of the two independent raters for the coding of the think-aloud interview  
11 transcripts was found to be significantly high ( $ICC_{\text{unjust}} = .99$ ;  $F_{439,439} = 143.15$ ;  $p < .001$ ).  
12  
13

#### 14 15 16 17 *More Quality Criteria of the PCK-items.* 18

19 For the internal consistency of the PCK-test scale "declarative knowledge" and the PCK  
20 component "pupils' errors", a very low Cronbach's alpha value ( $\alpha = .06$ ; 3 items;  $n = 65$ ) was  
21 found. For the two other knowledge dimensions, the Cronbach's alpha values was  $\alpha = .60$  and  
22  
23  
24  
25  
26  
27  $\alpha = .67$  ( $n = 65$ ) (see Table 5).  
28

29  
30  
31 [Insert Table\_5 about here]  
32  
33

34  
35  
36 Finally, the intraclass correlation coefficient (ICC) of the two independent raters for the open-  
37 ended PCK-items was found to be statistically significant ( $ICC_{\text{(unjust)}} = .80$ ;  $F_{588,588} = 5.10$ ;  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
 $p < .001$ ).

## 59 60 **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

1  
2  
3 and tenth grade German pupils reported in this article might be summed up as three main  
4 categories which might be transferred to other biological topics as well: *wrong locations of the*  
5 *structure; errors concerning the function* and *wrongly used terms and definitions*. The results of  
6 this study showed that most of the analysed pupils' error categories concerning the knee-jerk  
7 were usable for the categorisation of pupils' answers in this topic and that the interrater  
8 agreement (AC1) of the categories ranged from moderate to almost perfect (Gwet, 2002; Landis  
9 & Koch, 1977).

10  
11  
12 All in all, this study was able to show one way of developing PCK-items based on categorisation  
13 of pupils' errors which might enable researchers to develop their topic-specific, practical-  
14 orientated and empirically based PCK-items based on this idea. The quality of the developed  
15 PCK-items was validated by qualitative think-aloud interviews and the quantitative pilot phase.  
16 The results of these analyses demonstrated the development of a valid, reliable and objective  
17 PCK-test instrument for testing teachers' PCK concerning pupils' errors in biology.

18  
19  
20 The developed items in the PCK-test have curricular content validity because they were  
21 developed after curricular reviews and many meetings with experienced teachers. Additionally,  
22 the empirical data of the pupils' errors gave more content validity. Finally, the think-aloud  
23 interviews demonstrated that teachers mostly thought of or used the kind of knowledge  
24 (declarative, procedural or conditional) for answering the item which was intended to be  
25 measured by the respective item. The ICC of the interview transcript coding showed a  
26 satisfactory interrater agreement between the two independent raters (Field, 2009). In the end of  
27 all the interviews, four follow-up questions about the items were also asked. One of these  
28 questions asked the teacher if their answers to the items might represent their personal way of  
29 enactment in classroom situations or their way of thinking concerning example models in biology  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 lessons. This could be seen as face validity, because all the interviewed teachers said something  
4 similar to what Teacher K said: “My answers represent what I strive to do”.

5  
6  
7  
8 The internal consistency of the PCK-test (its Cronbach’s alpha values) showed satisfactory  
9 results except for the declarative knowledge of pupils’ errors—about teachers’ prediction of how  
10 often the given pupils’ error might arise after the learning process. One reason might be that  
11 teachers often wrongly predict their pupils’ learning outcomes and because they have over-  
12 expectations of their pupils’ learning outcomes (Helmke & Schrader, 1987; Hosenfeld, Helmke  
13 & Schrader, 2002).

14  
15  
16  
17 The objectivity of the PCK-items was also confirmed with an acceptable intraclass correlation  
18 coefficient (ICC).

19  
20  
21  
22 Due to the small sample size of this study for more detailed statistical analysis (e.g., factor  
23 analysis), further studies with samples of about 200 teachers will be conducted to test if all the  
24 developed items conform to the three dimensional theoretical model of the current project  
25 *ProwiN* (Tepner et al., sub.). Additionally, the items were translated into English and will be  
26 given to American teachers in our further studies. The results of the future think-aloud interviews  
27 might show the possible use and transferability of this study’s development ideas to other  
28 countries and cultures.

29  
30  
31  
32 Teachers’ knowledge about pupils’ possible errors—which may arise in a lesson or about  
33 possible reasons for these errors—is important for teachers because this knowledge has different  
34 advantages for teachers. On the one hand, it can be used as an orientation during the planning  
35 process of lessons (Schumacher, 2008; Seifried & Wuttke, 2010); and meaningful arrangement of  
36 materials, figures and models can also result. On the other hand, the information can be used for  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the training of pupils' error analysis; pupils have the possibility to learn the so-called *negative*  
4 *knowledge* (Oser, Hascher, & Spychiger, 1999). In the analysis of errors made by anonymous  
5  
6  
7  
8 pupils, the pupils have to change their point of view to that of a corrector. This change is the best  
9  
10 opportunity for learning from errors (Schumacher, 2008). Therefore, empirically analysed and  
11  
12 categorised pupils' errors, as reported in this article, are important for teachers. In addition, these  
13  
14 results can be used for test items construction in science education research as well as in teacher  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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, & Volkman, 2007). In research there are different results concerning the influence of teachers' 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

1  
2  
3 needed (Friedrichsen et al., 2007; Blömeke, Felbrich, & Müller, 2008). As well, biology teachers'  
4  
5 professional knowledge development (longitudinal) and the relationship between the  
6  
7 developments of the different components (PCK, CK, PK) could be focused on in the future to  
8  
9 close the gap for biology teachers' research (Cohen & Yarden, 2009; Abell, 2007).  
10  
11

## 12 13 14 15 16 17 **References**

- 18  
19 Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman  
20  
21 (Eds.), *Handbook of research on science education* (pp. 1105-1149). New Jersey: Lawrence  
22  
23 Erlbaum Associates.  
24  
25  
26  
27 Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in  
28  
29 learning and literacy talk about knowledge. *Review of Educational Research*, 61 (3), 315-  
30  
31 343.  
32  
33  
34 Allen, M. (2005). Encouraging pupil error may promote better understanding of a scientific  
35  
36 concept. *School Science Review*, 86, 79–85.  
37  
38  
39 Allen, M. (2010). Learner error, affectual stimulation, and conceptual change. *Journal of*  
40  
41 *Research in Science Teaching*, 47 (2), 151-173.  
42  
43  
44 Ashlock, R. B. (2005). *Error patterns in computation: Using error patterns to improve*  
45  
46 *instruction (9th Edition)*. New-York: Macmillan.  
47  
48  
49 Bainbridge, R. (1981). To err is human: Towards a more positive approach to young children's  
50  
51 mistakes in arithmetic. *Mathematics in School*, 10(5), 10–13.  
52  
53  
54 Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers'  
55  
56 mathematical knowledge, cognitive activation in the classroom, and student progress.  
57  
58 *American Educational Research Journal*, 47, 133-180.  
59  
60

- 1  
2  
3 Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content  
4 knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content*  
5 *knowledge* (pp. 147-161). Dordrecht: Kluwer.  
6  
7  
8  
9
- 10 Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des  
11  
12 erziehungswissenschaftlichen Wissens angehender Lehrkräfte [Assessment of student  
13 teachers' pedagogical knowledge]. In S. Blömeke, G. Kaiser & R. Lehmann (Eds.),  
14 *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen*  
15 *und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste*  
16 *Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 171–194). Münster, Germany:  
17 Waxmann.  
18  
19  
20  
21  
22  
23  
24  
25  
26
- 27 Bouvier, A.(1987). The right to make mistakes. *For the Learning of Mathematics*, 7(3), 17–25.  
28
- 29 Caleon, I. S., & Subramaniam, R. (2010). Do pupils know what they know and what they don't  
30 know? Using a Four-Tier diagnostic test to assess the nature of pupils' alternative  
31 conceptions. *Research Science Education*, 40, 313-337.  
32  
33  
34  
35
- 36 Chi, M. T. H., Slotta, J. D., & DeLeeuw, N. (1994). From things to processes: A theory of  
37 conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.  
38  
39  
40
- 41 Cokelez, A., Dumon, A., & Taber, K. S. (2008). Upper secondary French pupils, chemical  
42 transformations and the “Register of models”. *International Journal of Science Education*,  
43 30 (6), 807-836.  
44  
45  
46  
47
- 48 Cohen, R. & Yarden, A. (2009). Experienced Junior-High-School teachers' PCK in light of a  
49 curriculum change: "The cell is to be studied longitudinal". *Research of Science Education*,  
50 39, 131-155.  
51  
52  
53  
54
- 55 de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge.  
56  
57 *Educational Psychologist*, 31(2), 105 - 113.  
58  
59  
60

- 1  
2  
3 de Laughter, J. E., & Stein, S. (1998). Preconceptions about earth science among pupils in an  
4  
5 introductory course. *Eos*, 79, 429-434.  
6  
7  
8 Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory  
9  
10 and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual*  
11  
12 *change* (pp. 629-646). New York, London: Routledge.  
13  
14  
15 Driver, R.P. (1981). Pupils' alternative frameworks in science. *European Journal of Science*  
16  
17 *Education*, 3(1), 93-101.  
18  
19  
20 Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: The MIT Press.  
21  
22  
23 Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.  
24  
25 Fischer, H.E., Borowski, A., & Tepner, O. (in press). Professional knowledge of science teachers;  
26  
27 In B. Fraser; K. Tobin, & C. McRobbie (Eds.). *Second International Handbook of Science*  
28  
29 *Education*, New York: Springer.  
30  
31  
32 Friedrichsen, P., Abell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does  
33  
34 teaching experience matter? Examining biology teachers' prior knowledge for teaching in an  
35  
36 alternative certification program. *Journal of Research in Science Teaching*, 46, 357-383.  
37  
38  
39 Gilbert, J., Osborne, R., & Fensham, P. (1982). Children's science and its consequences for  
40  
41 teaching. *Science Education*, 66, 623-33.  
42  
43  
44 Gwet, K. L. (2008). *Interrater Reliability*. Wiley Encyclopedia of Clinical Trials, Copyright 2008  
45  
46 John Wiley & Sons, Inc.  
47  
48  
49 Gwet, K. L. (2002). Kappa Statistic is not satisfactory for assessing the extent of agreement  
50  
51 between raters. *Statistical Methods For Inter-Rater Reliability Assessment*, 1, 1-5. Retrieved  
52  
53 from [http://agreestat.com/research\\_papers/kappa\\_statistic\\_is\\_not\\_satisfactory.pdf](http://agreestat.com/research_papers/kappa_statistic_is_not_satisfactory.pdf) [Date  
54  
55 accessed: 25 May. 2011].  
56  
57  
58  
59  
60  
60 Hammann, M. (2003). Aus Fehlern lernen. [Learning by errors], *Unterricht Biologie*, 287, 31-35.

- 1  
2  
3 Hashweh, M. (2005). Teacher pedagogical constructions: A reconfiguration of PCK. *Teachers*  
4 *and Teaching: Theory and Practice*, 11, 273–292.  
5  
6  
7  
8 Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher  
9 judgment accuracy on achievement. *Teaching and Teacher Education*, 3, 91-98  
10  
11  
12 Helm, H. (1980). Misconceptions in physics amongst South African pupils. *Physics Education*,  
13 15, 92-7.  
14  
15  
16  
17 Henze, I., Driel van, J. H., & Verloop, N. (2008). Development of experienced science teachers'  
18 pedagogical content knowledge of models of the solar system and the universe. *International*  
19 *Journal of Science Education*, 30(10), 1321-1342.  
20  
21  
22  
23  
24 Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's  
25 mathematics professional development institutes. *Journal for Research in Mathematics*  
26 *Education*, 35 (5), 330-351.  
27  
28  
29  
30  
31 Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical  
32 knowledge for teaching on student achievement. *American Educational Research Journal*,  
33 42, 371-406.  
34  
35  
36  
37  
38 Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge:  
39 Conceptualizing and measuring teachers' topic-specific knowledge of pupils. *Journal for*  
40 *Reserach in Mathematics Education*, 39, 372-400.  
41  
42  
43  
44  
45 Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts-  
46 und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Lehr-  
47 Lern-Studie SALVE. [Diagnotic competency: teaching and learning relevant  
48 to student characteristics and the teachers' prediction in the teaching-learning study SALVE],  
49 In M. Prenzel & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische*  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 *Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen.*

4  
5 *Zeitschrift für Pädagogik*, 45. Beiheft (pp. 65-82). Beltz: Weinheim.

6  
7  
8 Houssart, J., & Weller, B. (1999). Identifying and dealing with misconceptions and errors in  
9  
10 primary mathematics: Student teachers record their experiences. *Mathematics Education*  
11  
12 *Review*, 11, 46-58.

13  
14  
15 Hsieh, H.-F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis.  
16  
17  
18 *Qualitative Health Research*, 15(9), 1277-1288.

19  
20 Jüttner, M., & Neuhaus, B. (2011). Vom Schülerfehler zum PCK – Test. Ansätze zur  
21  
22 Entwicklung von PCK – Items. [A PCK-test based on students' errors. Rudiments for  
23  
24 developing PCK items.] In U. Harms & I. Mackensen-Friedrichs (Eds.), *Lehr- und*  
25  
26 *Lernforschung in der Biologiedidaktik. Band 4* (pp. 27-39). Studien Verlag: Innsbruck.

27  
28  
29 Jüttner, M. & Neuhaus, B. (2010). Using empirically analyzed pupils' errors to develop a PCK-  
30  
31  
32 Test. In M.F. Taşar & G. Çakmakçı (Eds.), *Contemporary science education research:*  
33  
34 *preservice and inservice teacher education* (pp. 331-340). Ankara, Turkey: Pegem Akademi.

35  
36 Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential  
37  
38  
39 for progress. *Studies in Science Education*, 45(29), 169-204.

40  
41 Krüger, D. (2007). Die Conceptuale Change-Theorie. [The conceptual change theory]. In D.  
42  
43 Krüger & H. Vogt (Eds.), *Handbuch der Theorien in der biologiedidaktischen Forschung*  
44  
45 (pp. 81-92). Berlin: Springer Verlag.

46  
47  
48 Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical  
49  
50  
51 data. *Biometrics*, 33 (1), 159–174.

52  
53 Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of  
54  
55  
56 pedagogical content knowledge. *International Journal of Science Education*, 30, 1343-1363.

- 1  
2  
3 Leighton, J. P.(2009). Two Types of Think Aloud Interviews for Educational Measurement:  
4  
5 Protocol and Verbal Analysis. *Paper presented at the National Council on Measurement in*  
6  
7 *Education (NCME), CA.*  
8  
9
- 10 Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge  
11  
12 zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. [It depends on  
13  
14 the teacher. Empirical evidences of the connection between teachers' competence, teachers'  
15  
16 acting and the learning of the pupils], *Zeitschrift für Pädagogik*, 52 (51), 47-70.  
17  
18
- 19 Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in  
20  
21 science teacher education. *International Journal of Science Education*, 30 (10), 1301-1320.  
22  
23
- 24 Magnusson, S., & Krajcik, J. & Borko, H. (1999). Nature, sources and development of  
25  
26 pedagogical content knowledge for science teaching. In: Gess-Newsome & Lederman (Eds.),  
27  
28 *Examining pedagogical content knowledge: The construct and its implications for science*  
29  
30 *education* (pp. 95-131). Dordrecht. Kluwer Academic Publishers.  
31  
32
- 33 Mayring, P. (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung / Forum:  
34  
35 Qualitative Social Research, North America, 1. Available at: [http://www.qualitative-](http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385)  
36  
37 [research.net/index.php/fqs/article/view/1089/2385](http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385). [Date accessed: 26 Oct. 2010].  
38  
39
- 40 Nebah, S. J. (2002). *Do people learn from mistakes? A pedagogical study and analysis of*  
41  
42 *autobiographies and interviews*. Dissertation am Departement Erziehungswissenschaften der  
43  
44 Universität Fribourg.  
45  
46
- 47 Novak, G.S. (1977). *Representations of knowledge in a program for solving physics problems.*  
48  
49 *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.*  
50  
51 Cambridge, MA: MIT Press.  
52  
53
- 54 Oser, F., Hascher, T., & Spsychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie der  
55  
56 „negativen“ Wissens. [Learning out of errors. About the psychology of the "negative"  
57  
58  
59  
60



1  
2  
3 knowledge]. In W. Althof (Ed.), *Fehlerwelten: vom Fehlermachen und Lernen aus Fehlern*  
4  
5 (pp. 11-43). Opladen: Leske + Budrich.

6  
7  
8 Oser, F., & Spsychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens*  
9  
10 *und zur Praxis der Fehlerkultur*. [Learning is painful. About the theory of the negative  
11  
12 knowledge and about the practice of the culture of errors]. Weinheim: Beltz.

13  
14  
15 Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary*  
16  
17 *Educational Psychology*, 8, 293-316.

18  
19  
20 Park, S., Jang, J.-Y., & Chen, Y.-C. (2010). Is Pedagogical Content Knowledge (PCK) necessary  
21  
22 for reformed science teaching?: Evidence from an empirical study. *Research in Science*  
23  
24 *Education*, 41, 245–260. doi: 10.1007/s11165-009-9163-8

25  
26  
27 Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of Pedagogical Content  
28  
29 Knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals.  
30  
31 *Research in Science Education*, 38, 261-284.

32  
33  
34 Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Measuring declarative and  
35  
36 reflective components of biology teachers' pedagogical content knowledge. In M. F. Tasar &  
37  
38 G. Cakmakci (Eds.), *Contemporary science education research: Teaching* (pp. 71-77).  
39  
40 Ankara, Turkey: Pegem Akademi.

41  
42  
43 Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L. et al. (2007). *The*  
44  
45 *Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries.*  
46  
47 *Mathematics Teaching in the 21st Century (MT21)* (Rep. No. MSU Center for Research in  
48  
49 Mathematics and Science Education).

50  
51  
52  
53 Schori, E., Kersten, B. & Abderhalden, C. (2006). Wie reliabel ist der Fragebogen zur Analyse  
54  
55 der Pflegedokumentation IzEP-Dok© im Instrument zur Erfassung von Pflegesystemen  
56  
57 IzEP©? [How Reliable is the Questionnaire for the Analysis of Nursing Documentation  
58  
59  
60

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 Fehlerkultur. [An open play. Learning by errors and the development of the culture of errors]. In R.

1  
2  
3 Caspary (Ed.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (pp.  
4  
5 25-48). Freiburg: Herder.

6  
7  
8 Taber, K. S. (2009). *Progressing science education: Constructing the scientific research*  
9  
10 *programme into the contingent nature of learning science*. Dordrecht: Springer.

11  
12  
13 Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: from a scholarly  
14  
15 approach to teaching to the scholarship of teaching. *Cell Biology Education*, 4, 1-6.

16  
17  
18 Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher*  
19  
20 *Education and Development Study in Mathematics (TEDS-M): Conceptual framework*. East  
21  
22 Lansing, MI: Teacher Education and Development International Study Center, College of  
23  
24 Education, Michigan State University.

25  
26  
27 Tepner, O., Borowski, A., Fischer, H.E., Jüttner, M., Kirschner, S., Leutner, D. et al. (submitted).  
28  
29 Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von  
30  
31 Lehrkräften in den Naturwissenschaften. [Theoretical model for the development of test-  
32  
33 items to collect data about professional knowledge of science teachers], *Zeitschrift der*  
34  
35 *Naturwissenschaften*.

36  
37  
38 Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students'  
39  
40 misconceptions in science. *International Journal of Science Education*, 10 (2), 159-169.

41  
42  
43 Tsai, C.-C., & Chou, C. (2002). Diagnosing pupils' alternative conceptions in science. *Journal of*  
44  
45 *Computer Assisted Learning*, 18, 157-165.

46  
47  
48 Tsamir, P., Rasslan, S., & Dreyfus, T. (2006). Prospective teachers' reactions to Right-or-Wrong  
49  
50 tasks: The case of derivatives of absolute value functions. *Journal of Mathematical*  
51  
52 *Behavior*, 25, 240-251.

53  
54  
55 van Driel, J. H., Verloop, N., & Vos, W. d. (1998). Developing science teachers' Pedagogical  
56  
57 Content Knowledge. *Journal of Research in Science Teaching*, 35, 673-695.

1  
2  
3 van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in  
4  
5 science education: The role of teachers' practical knowledge. *Journal of Research in Science*  
6  
7  
8 *Teaching*, 38, 137-158.  
9

10  
11  
12  
13  
14  
15 \* the presented project is funded by the Federal Ministry of Education and Research; it is a cooperation project embedded in the  
16  
17 framework program of 'empirical research in education' (01JH0904).  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1

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

Term	Exemplary Literature
<a href="#">conceptions or misconceptions</a>	Allen, 2010; Chi, Slotta, & de Leewu, 1994;
	Duit, Treagust, & Widodo, 2008; Helm, 1980;
	Krüger, 2007 ; Treagust, 1988;
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; <a href="#">Seifried &amp;</a>
	<a href="#">Wuttke, 2010</a>
<a href="#">children's science</a>	Gilbert et al., 1982

Deleted: (mis)conception

Deleted: C

Deleted: '

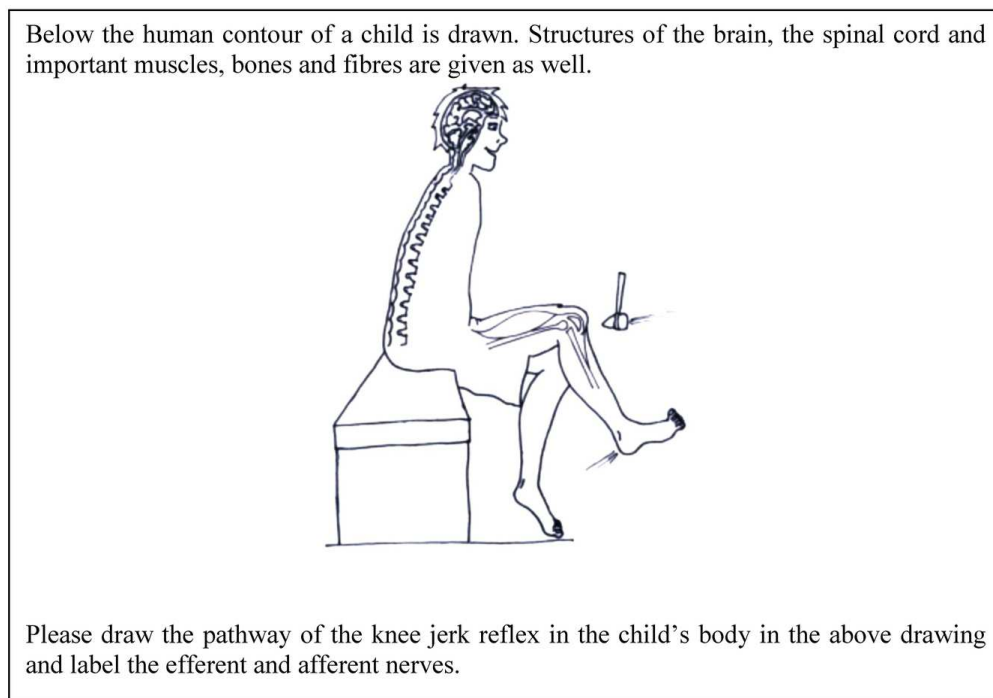


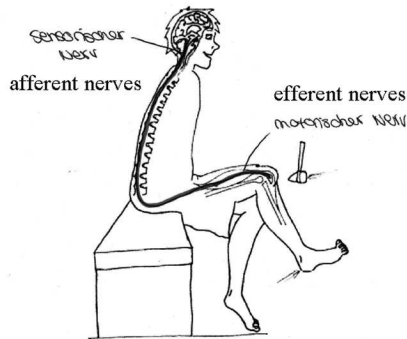
Figure 1.

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)

new Only

In a students' achievement test, the students in the 9th grade had to draw the pathway of a knee-jerk into a given contour of a child by labelling the efferent and afferent nerves. In the human body some structures were given: the brain, the spinal cord and important muscles, bones and fibres. At the time of the test, the students had already learned about this topic. Shown below is one exemplary answer of a student:



*Students' error:*

*The drawn impulse reaches the brain. Only the afferent nerve is drawn into the body. Additionally the starting point of this nerve has a wrong location (it starts at the patella and not in the muscle). Finally the labelling is missing completely.*

- a) Please try to determine how often the following error could arise when students of your school type had learned about the topic and then had to fill in the following answer.

*"The drawn impulse reaches the brain."* \_\_\_\_\_ %

- b) Please describe as many reasons as possible for the following error:

*"Beginning and/or ending of the nerve fibres are not located at the muscle, but in front of the patella."*

Therefore you should think of material that could be used in lessons or conceptions that could arise out of everyday life.

- c) Situation:

After the students in the 9th grade had learned everything about the human central nervous system and its functions, they took an achievement test in class. In this test the students were asked to draw the reflex arc of the knee-jerk on a drawing of a given human body. Additionally they had to label the afferent and efferent nerves. One student drew only one pathway of the impulse and therefore he named one end afferent and the other efferent (see the figure above).

Please describe all the thinkable possibilities of how you could react to this error in this specific situation.

Figure 2.

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)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2

Deleted: 3

Summary of Seven Identified Categories of Typical Students' Errors (n = 461) about Knee-Jerk (cf. Jüttner & Neuhaus, 2011)

Deleted: authors, in press

Deleted: student

Identified categories of pupils' errors	frequency
1.) the drawn impulse reaches the brain	63%
2.) beginning and/or ending of the nerve fibres are not located at the muscle, but...	
a. in front of the patella	47%
b. at bones	42%
c. in the heel or at the tibia	14%
d. on the tendon	12%
3.) there is only one way for the impulse illustrated by one single nerve	34%
4.) the clearing up centre of the nerves is located in the brain	26%
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	15%
7.) nerves are not named according to their function, but to their absolute location (brain = afferent nerve; muscle = efferent nerve)	7%



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

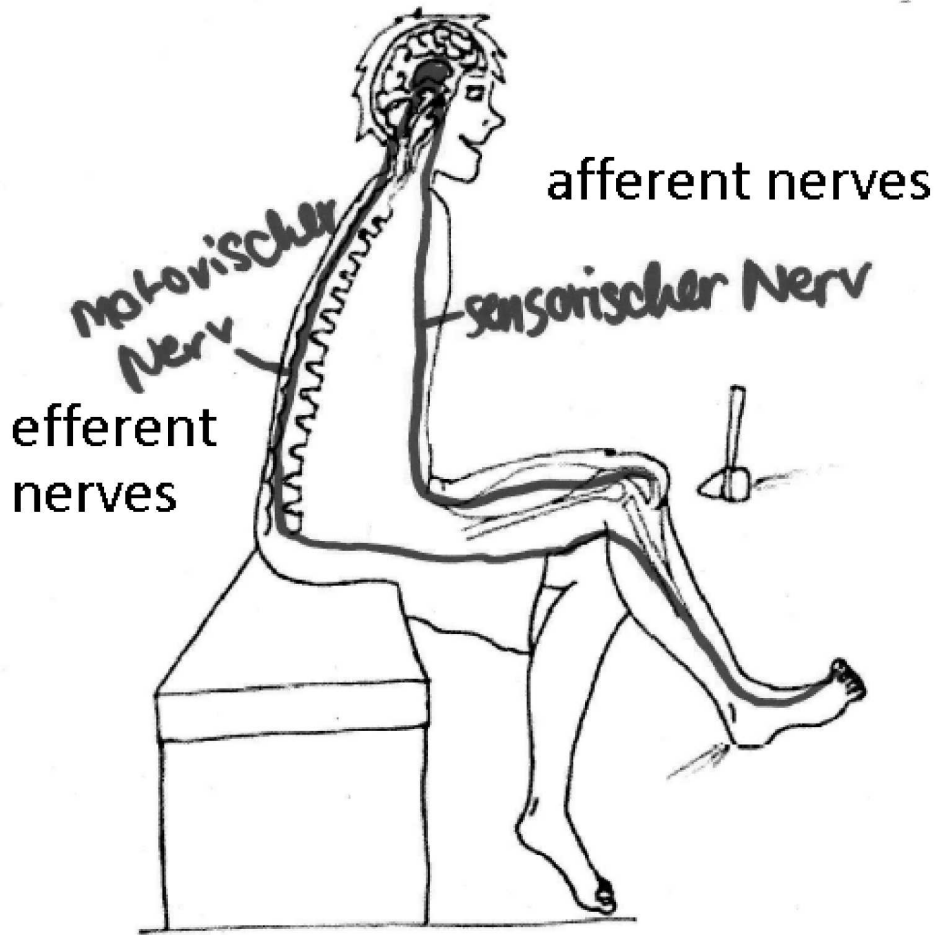


Figure 3.  
Example of a pupil's answer to the task on the reflex arc (Figure 1). Category 1 as well as category 2b, 2c, 4 and 5 are identified here.  
136x133mm (600 x 600 DPI)

mu

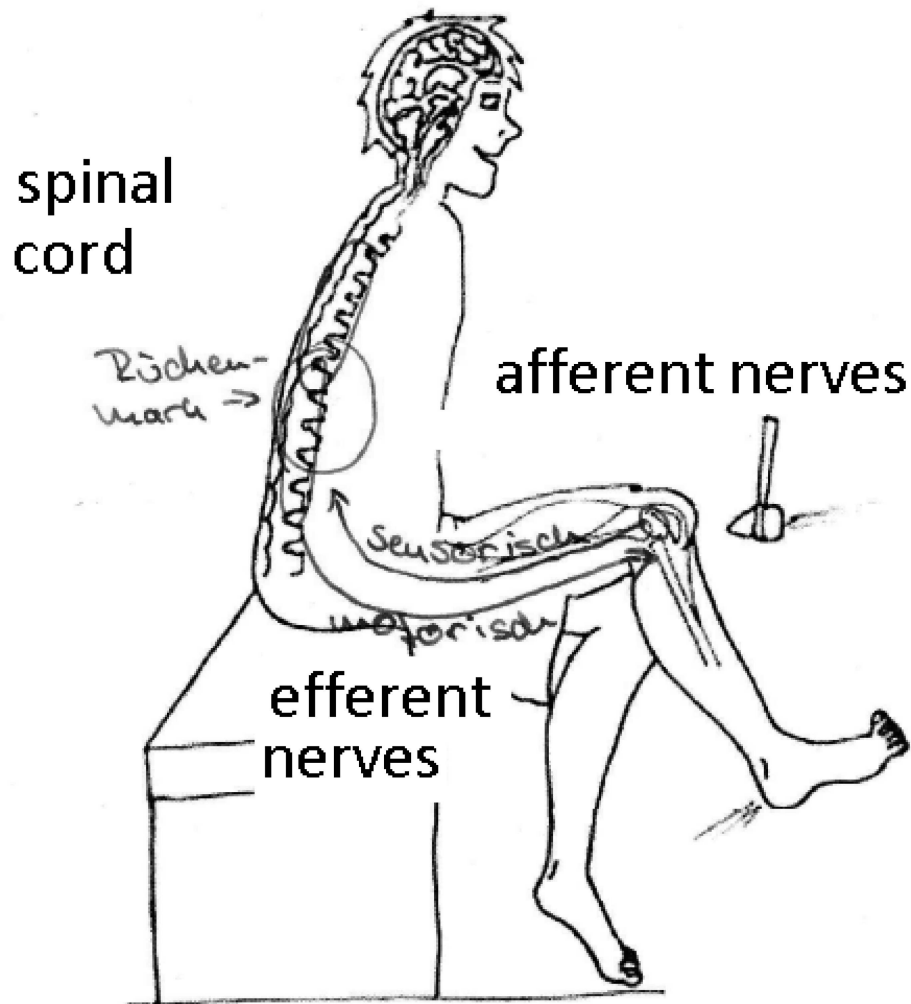


Figure 4.

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

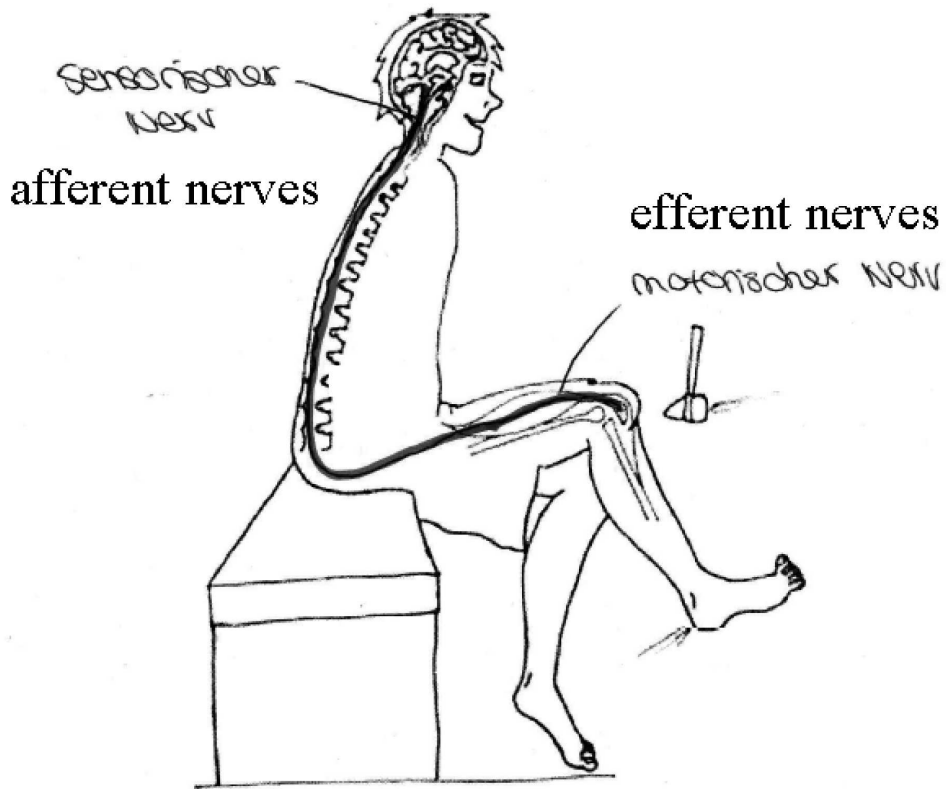


Figure 5.  
The sample of a pupil's answer to the item seen in Figure 1. Category 1, 2a, 3 and 7 could be identified here.  
162x135mm (600 x 600 DPI)

Only

Table 3

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

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

Table 4Means and Standard Deviations for the First Coder's Coded Categories of Teachers' (n = 5)Think-aloud Interview Transcripts Concerning Developed PCK-items (cf. Figure 2)

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

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

Table 5.

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

knowledge dimension	facet	items	Cronbach's alpha
declarative knowledge	pupils' errors	3	$\alpha = .06$
procedural knowledge	pupils' errors	3	$\alpha = .60$
conditional knowledge	pupils' errors	6	$\alpha = .67$
<hr/>			
all knowledge dimensions	pupils' errors	12	$\alpha = .65$

- Deleted: 4
- Deleted: c
- Deleted: d
- Deleted:
- Deleted: d
- Deleted: f
- Deleted: p
- 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