Educ. quím., **25**(3), 380-390, 2014. © Universidad Nacional Autónoma de México, ISSN 0187-893-X
Publicado en línea el 11 de junio de 2014, ISSNE 1870-8404

Development of a Pedagogical Content Knowledge test of chemistry language and models

Martina Strübe,[1] Holger Tröger,[2] Oliver Tepner,[2] and Elke Sumfleth[1]

ABSTRACT

In 1986 Shulman developed a typology of teachers' professional knowledge. Since this, research on teachers' professional knowledge, especially on the measurement of professional knowledge, has increased. Measuring teachers' professional knowledge requires tests which focus on specific knowledge types and subjects. However, there are only few professional knowledge tests analysing teachers' pedagogical content knowledge of using models and chemistry language in chemistry classes. The following article describes the development of a pedagogical content knowledge test for chemistry teachers, which focuses on teachers' pedagogical content knowledge regarding the handling of models and chemistry language. As a result the test measures the intended construct reliably.

KEYWORDS: pedagogical content knowledge, models, chemistry language

Resumen (Desarrollo de una prueba de Conocimiento Pedagógico del Contenido sobre modelos y el lenguaje de la química)

En 1986 Shulman desarrolló una tipología del conocimiento profesional de profesores. Desde ese momento, la investigación sobre el mismo se ha incrementado. La medición del conocimiento profesional de profesores requiere de pruebas que se enfoquen sobre tipos y contenidos específicos. Sin embargo, existen solamente unas cuantas pruebas de conocimiento profesional que analicen el conocimiento pedagógico del contenido sobre el uso de modelos y el lenguaje de la química en las clases. El siguiente artículo describe el desarrollo de una prueba sobre conocimiento pedagógico del contenido para profesores de química sobre este tópico en particular. Como resultado, la prueba parece medir el constructo pretendido fidedignamente.

Palabras clave: conocimiento pedagógico del contenido, modelos, lenguaje de la química

Introduction

Teachers' professional competence is supposed to have an influence on students' achievement (Kunter, Kleickmann, Klusmann, and Richter, 2011), and includes for example teachers' motivation, beliefs and professional knowledge (Baumert and Kunter, 2006). Teachers' professional knowledge is a substantive precondition for their competent acting in classroom situations. In the last centuries, different national and international studies have already measured and analysed teachers' professional knowledge (e.g. COACTIV, MT21). In Germany, research on pedagogical content knowledge in chemistry, especially on using models and chemistry language in class, is limited. Analysing teachers' pedagogical content knowledge focusing on using models and chemistry language, especially in large-scale assessments, requires a valid, reliable, and objective test-instrument.

Teachers' Professional Knowledge

Shulman (1986, 1987) describes seven types of teachers' professional knowledge ("curriculum knowledge", "knowledge of educational ends, purposes, and values, and their philosophical and historic grounds", "general pedagogical knowledge", "content knowledge", "pedagogical content knowledge", "knowledge of learners and their characteristics", and "knowledge of educational contexts" (Shulman, 1987, p. 8)). Contemporary research mainly focuses on content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) (Baumert and Kunter, 2006).

Pedagogical content knowledge is described by Shulman (1987) as a "special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding" (Shulman, 1987, p. 8). At a PCK summit in 2012 a definition of pedagogical content knowledge was devised by a workgroup led by Gess-Newsome, Carlson, and Gardner. They defined pedagogical content knowledge as "the knowledge of, reasoning behind,

[1] University of Duisburg-Essen, Chemistry Education, Schuetzenbahn 70, 45127 Essen, Germany.
E-mails: martina.struebe@uni-due.de, hoelger.troger@uni-due.de, elke.sumfleth@uni-due.de
[2] University of Regensburg, Chemistry Education, Universitaetsstraße 31, 93040 Regensburg, Germany.
E-mail: oliver.tepner@chemie.uni-regensburg.de

planning for, and enactment of teaching a particular topic in a particular way for a particular reason to particular students for enhanced student outcomes" (Gess-Newsome, 2013; Garritz, 2013). In this article pedagogical content knowledge is defined as the knowledge that enables teachers to structure, link, represent, and explain the content to students (Schmelzing, Wüsten, Sandmann, and Neuhaus, 2010; Krauss, Neubrand et al., 2008). This includes the knowledge of how to present the content of a subject comprehensibly to learners by using e.g. analogies and demonstrations (Shulman, 1986). In addition, pedagogical content knowledge involves knowledge about students' conceptions and misconceptions, and how to deal with them (Shulman, 1986; Garritz, 2013).

Shulman (1987) describes pedagogical content knowledge as an amalgam of content knowledge and pedagogical knowledge. Based upon this assumption, a correlation between the two knowledge categories should be expected. The correlation between pedagogical content knowledge and content knowledge of mathematic teachers has been examined by Krauss, Brunner et al. (2008). They concluded that teachers who taught mathematics at *Gymnasium*[1] (GY) scored higher in content knowledge and pedagogical content knowledge than mathematic teachers of other secondary schools. However, they could not "distinguish the two knowledge categories empirically in the high-expertise group of GY teachers, but that this distinction was clearly visible in the group of NGY[2] teachers" (p. 724). Differences between types of school could be found in chemistry as well (Tepner and Dollny, 2014). Chemistry teachers teaching at the GY score higher at the content knowledge (CK) test and pedagogical content knowledge (PCK) test than teachers of other secondary schools. However, reported differences in PCK regarding different types of school are smaller if effect of CK variance on PCK variance is controlled by including CK as a covariate (Tepner and Dollny, 2014). Overall, content knowledge seems to be a precondition for developing pedagogical content knowledge (Baumert et al., 2010; Tepner and Dollny, 2014).

[1] In Germany, after primary school students and parents can choose between four types of secondary school in Germany. The difference between these four is the intensity of general education. In the lower secondary schools (Realschule and Hauptschule), general education is not as intense as in secondary schools (comprehensive school and Gymnasium) (Tepner and Dollny, 2014). Students finish lower secondary school at about the age of 16 years. They are qualified to do an apprenticeship e.g. as a cook, mechanic or administration officer. Students who get a "high school graduation" (German Abitur), finish comprehensive school (Gesamtschule) at the age of 19 years or Gymnasium at the age of 18 years. These students have the opportunity to study at university.

[2] Annotation of the authors: NGY means non-Gymnasium and is synonymous with lower secondary schools.

The Facets Models and Chemistry Language

Using Models in Class

Models play an important role in the acquirement of knowledge in science and science education (Justi and Gilbert, 2002a). "(...), they function as a bridge between scientific theory and the world-as-experienced ('reality')" (Gilbert, 2004, p. 1169). Based on Hodson's (as cited in Justi and Gilbert, 2002a) purposes for science education ("learning of science", "learning about science", "learning to do science"), Justi and Gilbert (2002a, 2003) describe the role of models in science education: Students should know the most important models in science, how they were developed and the limitations of models. They should develop and test their own models and know about the importance of models when scientific findings were disseminated and accepted (Justi and Gilbert, 2002a; 2002b; 2003). In this context Gilbert (2004) speaks of "Learning to Use Models", "Learning to Revise Models", and "Learning the Reconstruction of a Model". These intentions can be found in national and international standards which ask for using models, developing, and testing models and reflecting on models in class (NRC, 1996; KMK, 2005).

Teachers can help students learning about and with models by differentiating between the model and the experience (Mikelskis-Seifert, 2009; Saari and Viiri, 2003). In addition, it is important to discuss the limitations of models (Justi and van Driel, 2005; Saari and Viiri, 2003) and to carve out the change or replacement of models (Maia and Justi, 2009; Mikelskis-Seifert, 2009). It is also necessary for learning to use different models which represent a concept under different aspects or for different purposes (Grosslight, Unger, and Jay, 1991; Harrison and Treagust, 2000; Saari and Viiri, 2003). The colour of a model can lead to students' misconceptions, because of this it is important to discuss the function of the colour in class (Justi and van Driel, 2005). Teachers should involve students in modelling processes, by creating, developing, building, testing, communicating, and reflecting their own models (Gilbert, 2004; Grosslight et al., 1991; Henze, Van Driel, and Verloop, 2007a; Justi and Gilbert, 2003; Maia and Justi, 2009).

In order to do so, it is important for teachers to know how to create learning opportunities which include e.g. adequate teaching models, modelling activities and reflection on models (Gilbert, 2004; Henze, Van Driel, and Verloop, 2007a; Justi and van Driel, 2005). Recent studies indicate a small teachers' knowledge about models and modelling in science (Henze, Van Driel, and Verloop, 2007b; Justi and Gilbert, 2002a, Justi and van Driel, 2005). Research on teachers' pedagogical content knowledge on using models and modelling is rare in Germany.

Operationalization of Chemistry Language

Before discussing the importance of language and chemistry language in class, the meaning of language and communication is reflected.

Communication is the basis for human interaction. However, defining communication *per se* is not possible. A linguistic approximation to the construct *communication* describes it as any form of information processing between humans or data processing machines and humans by using signs and symbols (Bußmann, 2002). Communication comprises the use of body language, non-verbal elements, as well as talking and presenting orally or written in front of an audience. In addition, illustrating and presenting facts with e.g. graphics, as well as intrapersonal (e.g. monolog, affirmation) and interpersonal abilities (e.g. asking, listening, feedback) are part of communication. Transferring information is the basis of every form of communicating (Eunson, 2012). Models of communication describe modelling information transfer. Communicational models always depend on their context of development and scope of application and usually contain a sender and a recipient which are a communicational unit and are related in a certain way (Bußmann, 2002; Eunson, 2012; Grucza, 2012; Kessel and Reimann, 2008).

Language is a characteristic of communication (Eunson, 2012) and is used for exchanging thoughts, beliefs, and knowledge (Bußmann, 2002). From a linguistic point of view, language is the primary medium of communication. Although definitions of language are very different, a general definition describes languages as distinct systems of arbitrary but conventionally conveyed and used signs. Language is central for the exchange of and about knowledge (Bußmann, 2002).

For the exchange on a subject related level, technical languages are used (Grucza, 2012; Rincke, 2007; Roelcke, 2005). Research of different disciplines gives attention to the development and use of technical language as well as finding ways to foster students' use of technical language (Becker-Mrotzek, 2013; Buhlmann and Fearns, 2000). Although, current research misses a consistent definition of technical language, different approaches emphasize common characteristics of technical language that can be used for forming a definition. The central characteristic of a technical language is its' functional fixedness: It is used for the effective communication between experts of a specific domain. Technical language is shaped by the use of a subject specific lexis (technical terms) (Fluck, 1996; Grucza, 2012; Rincke, 2007; Wellington and Osborne, 2001) which should be free of everyday connotations (Bußmann, 2002). Besides, technical language is characterized by complex syntactic structures, nominalizations, and compositions (Schmölzer-Eibinger, 2013) as can be found in written communication (Koch and Oesterreicher, 1985, 2007). Based upon these definitions, technical languages are subject- and content- specific and are used for communicating about these contents. Especially in class, language is fundamental for teaching and learning (Norris and Phillips, 2003; Wellington and Osborne, 2001; Yore and Treagust, 2006). Language, especially technical language, is deeply connected to the learning in every subject (Merzyn, 2008; Özcan, 2013; Schmölzer-Eibinger, 2013; Sumfleth and Pitton, 1998). Technical language links students' everyday life and science (Parchmann and Bernholt, 2013). Characteristics of technical languages are e.g. a complex syntax and a formal conception (Koch and Oesterreicher, 2007; Schmölzer-Eibinger, 2013). In contrast to this, subject specific lexis is not generalised across different subjects. In chemistry, learning content is correlated to students' skills with chemistry language. Poor chemistry language skills are associated with difficulties in learning chemistry (Özcan, 2013). In conclusion, facilitating students' technical language is desirable for subjects like chemistry and physics (Becker-Mrotzek, 2013; Buhlmann and Fearns, 2000; Busch and Ralle, 2013; Kulgemeyer and Schecker, 2013; Özcan, 2013; Sumfleth, Kobow, Tunali, and Walpuski, 2013; Vollmer and Thürmann, 2013). Furthermore, teachers should use technical language in a sophisticated way, in order to foster students' use of technical language in class (Linneweber-Lammerskitten, 2013; Schmölzer-Eibinger, 2013; Vollmer and Thürmann, 2013).

Development of the PCK Test

Measuring Teachers' Professional Knowledge

Several studies use interviews, questionnaires, tests, video analyzes or mixed methods approaches (e.g. Henze, Van Driel, and Verloop, 2007a; Justi and Gilbert, 2003; Krauss, Baumert, and Blum, 2008; MT21; Riese and Reinhold, 2008; Van Driel, de Jong, and Verloop, 2002) to analyze and measure teachers' professional knowledge.

The subject specific professional knowledge test of Dollny (2011) measures content knowledge (e.g. about the periodic table of elements, acids and bases) and pedagogical content knowledge of chemistry teachers. The content knowledge items are designed in a multiple-choice single-select format. The pedagogical content knowledge items have a closed-item format with a Likert-scale, ranging from 1 (e.g. "very important", "very meaningful") to 6 (e.g. "unimportant", "meaningless") (Dollny, 2011; Witner and Tepner, 2011). The items were constructed using the "Item Development Model for Assessing Professional Knowledge of Science Teachers" (Tepner et al., 2012) (Fig. 1). The three axes of the model describe the knowledge areas, the PCK-facets and the content. The knowledge axis is divided into procedural, declarative, and conditional knowledge. Conditional knowledge composes the knowledge of when and why an activity or process is executed. Procedural knowledge is the knowledge about how behaviour and processes in class are created. Facts and terms form the declarative knowledge. Originally, the PCK-facets included the knowledge of experiments, students' conceptions, and models as central elements of PCK (Tepner at al., 2012; cf. Jüttner, Boone, Park, and Neuhaus, 2013; Tepner and Dollny, 2014; Van Driel, de Jong, and Verloop, 2002). Later on, teachers' handling of chemistry language was added to the facets.

Although the PCK test of Dollny (2011) contains some items referring to conditional and declarative knowledge of

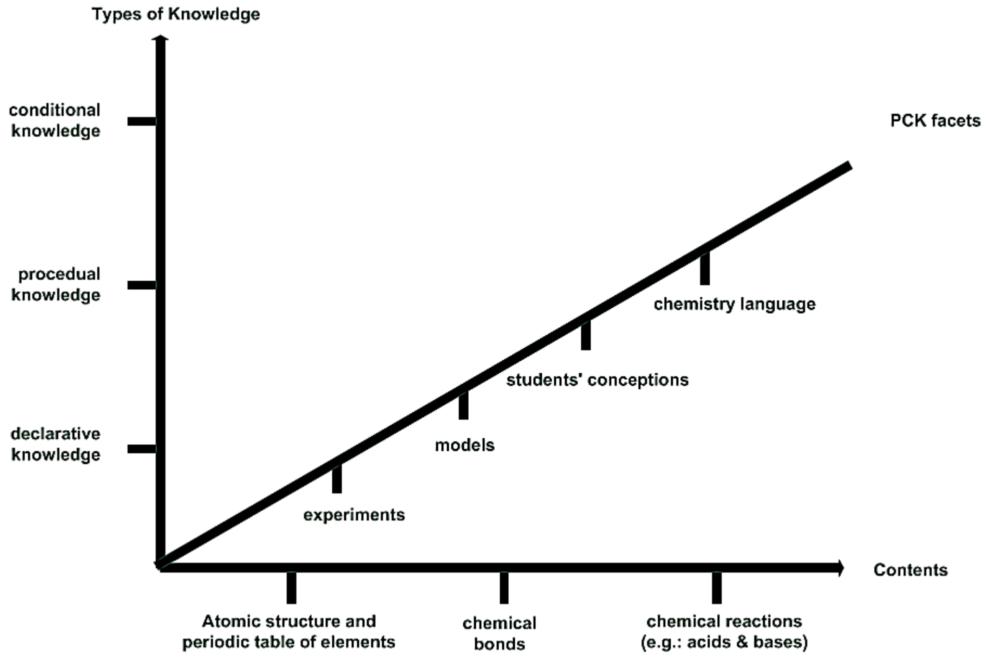

Figure 1. "Item Development Model for Assessing Professional Knowledge of Science Teachers" adapted of Tepner et al., 2012.

models, there were not enough items about models and no items measuring the handling of chemistry language in class. For this reason, new PCK items focusing on using models and chemistry language in class have been developed. The items are included in a new PCK test called FEMo (German short form for chemistry language (*Fachsprache*), experiments and models) and were developed by the authors. Items which focus on the use of experiments in class are added for the main study and have been developed by Tepner, Backes and Sumfleth (unpublished). Those items are not part of this article.

The PCK items concerning the use of models and chemistry language were evaluated in a pilot study. Based upon an expert-rating a point of reference was developed. Two booklets, each containing 15 items (a set of PCK items of using models and chemistry language) were developed for the pilot study. All items are in a closed item-format, so teachers can rate on a 6-point Likert-scale.

Development of Pedagogical Content Knowledge Model Items

The structure of the pedagogical content knowledge item is in line with the structure of the ones developed by Dollny (2011). Each item has an item stem that describes a classroom situation, e.g. a presentation of a student drawing which represents a model or a model which should be used. Every situation is presented as if the teacher, who fills out the test, helps a novice teacher. An item has four answers which describe a possible reaction or behaviour in this situation. These answer alternatives were developed theoretically and include the research findings and demands how to deal with models in class as described above. The items are subdivided into modelling processes; criticizing models, knowledge of models, and use of models. Items labelled use of models deal—e.g. with the adequacy of a given model for explaining a modelled concept. Items on criticizing models focus on—e.g. how a teacher could create instructions or learning opportunities which deals with e.g. the limitations of a given model (scientific model (Bohr) or real model (sodium chloride cube model)). The items on knowledge of models include e.g. dealing with different models, modelling the same phenomenon or the "upgrade" of models (e.g. Dalton's atomic theory to Rutherford's atomic theory).

Initially, more than thirty items were developed while fifteen items were selected for the pilot study. Two of these have no Likert-scale. Teachers have to rank the answers, because one item describes a process and the other one involves the structure of acquiring knowledge by using models. For these two items just one answer is definitely right. The other items are rated on a Likert-scale from 1 ("true", "very good suitability", "very important") to 6 ("not true", "inadequate suitability", "not important").

Development of Pedagogical Content Knowledge Items of using Chemistry Language

The subject specific professional knowledge test of Dollny (2011) implicitly tests the teachers' knowledge of chemistry language. The knowledge of chemistry language is intrinsically tied to content knowledge (Merzyn, 2008). While knowledge of chemistry language is not a part of pedagogical content knowledge, the knowledge of using and handling of chemistry language is an aspect of pedagogical content knowledge (Riese and Reinhold, 2012). For this reason, it was necessary to develop items for the new test which gather the knowledge of using and handling chemistry language and not the knowledge of chemistry language.

In line with the structure developed by Dollny (2011) is the structure of the items regarding the knowledge of using

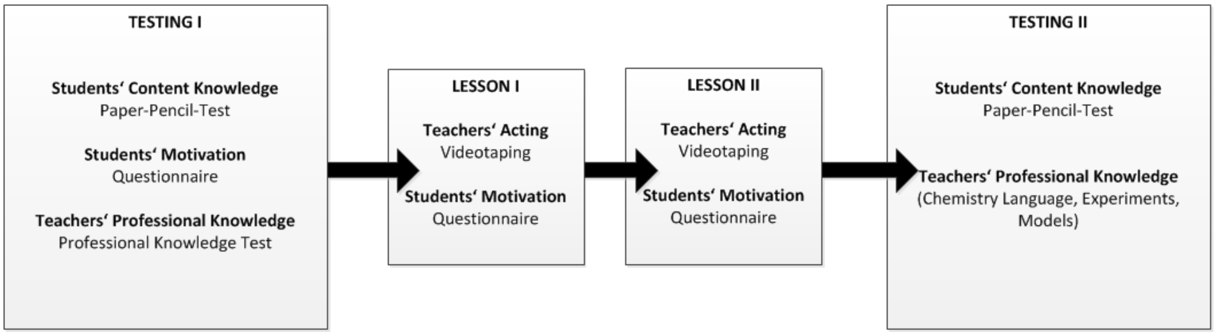

Figure 2. Design of the video study of the ProwiN project.

chemistry language. In the beginning of every item, the situation in class is briefly described. Following the description, a short dialog between a pre-service teacher and one or more students is presented. The dialogs are inspired by videotaped lessons of a previous study in a comparable grade (Pollender, unpublished). Four answer alternatives or statements which focus on teachers' behaviour in the dialog or possible activities in class are given in the test. Teachers have to rate these answers or statements on a Likert-scale ranging from 1 ("excellent", "applicable") to 6 ("inadequately", "not applicable"). This scale refers to German school marks (1 ("excellent") to 6 ("inadequately") and is familiar to teachers.

15 items with four possible activities each were constructed as described above.

The Study

Project Setting

The newly developed PCK test FEMo is part of the project ProwiN (Professional Knowledge in Science (*Professionswissen in den Naturwissenschaften*) which is founded by the German Ministry of Research and Education. ProwiN is subdivided into two phases. In the first phase, three teachers' subject specific professional knowledge tests focusing on teachers' CK and PCK in biology, chemistry, and physics have been developed and evaluated (Dollny, 2011; Jüttner, 2013; Kirschner, 2013). These tests were supplemented by a pedagogical knowledge (PK) test (Thillmann and Wirth, unpublished).

In the second phase, a video study on chemistry teachers' professional knowledge analyses the relationship between chemistry teachers' professional knowledge, their behaviour in class, and students' achievement (Troeger, Struebe, Sumfleth, and Tepner, 2014). Two lessons of the unit atomic structure and periodic table of elements are videotaped in this study (Fig. 2). These lessons will be analysed with coding-manuals focusing on the use of experiments, models, chemistry language, and dealing with students' conceptions. In a pre/post test design, students' achievement, interest, and motivation are measured by several paper-pencil-tests (e.g. knowledge about atomic structure and periodic table of the elements). Teachers' professional knowledge is measured by two paper-pencil-tests (Troeger et al., 2014). At the beginning of the unit, teachers process the professional knowledge test developed by Dollny (2011), and Thillmann and Wirth (unpublished). The FEMo is processed in the end of the unit.

Pilot Study and Expert-Rating

In summer 2012 a pilot study and an expert rating were conducted. In the pilot study teachers of an advanced teacher training were asked to fill out the new test. In addition, several inservice teachers did an online survey. The items were distributed via two booklets. Each booklet contained fifteen items. Booklet A contained seven PCK items on models and eight on chemistry language. Booklet B contained seven PCK items on chemistry language and eight items referring to models. Every teacher filled out one booklet (booklet B: N = 23 (14 females, 9 males); age mean value = 36.8; SD = 7.938; booklet A: N = 26 (16 females, 10 males); age mean value = 40.0; SD = 11.79), see also Tab. 1, 2, 3). The testing took 30 minutes. Teachers who used the online survey got a link for one booklet.

Nine German professors of chemistry education at six

Table 1. Teaching background of the sample (Booklet B).

	M	*SD*	Min	Max
Years of service	9.38	8.78	1	40
Years of teaching chemistry	9.167	8.89	1	40
Teaching chemistry hours per week	10.0	5.5	0	20

Table 2. Teaching background of the sample (Booklet A).

	M	*SD*	Min	Max
Years of service	13.54	11.26	2	38
Years teaching chemistry	13.92	11.44	2	38
Teaching chemistry hours per week	10.12	5.12	0	18

M = Mean value, *SD* = standard deviation, Min = Minimum, Max = Maximum. Years of service means the years in which a teacher is in service. Years of teaching chemistry could differ from years of service, because teachers can acquire an additional qualification for teaching chemistry in an advanced teacher training.

Table 3. Distribution of teachers working at different types of school.

	Realschule	*Gymnasium*	Comprehensive school
$n_{\text{teachers of booklet A}}$	1	22	3
$n_{\text{teachers of booklet B}}$	2	16	5

universities (2 female, 6 males) and one teacher trainer (1 male) conducted the additional expert rating. The two booklets were either sent to the experts or filled out online. The aim of the expert rating is to create a point of reference for appraising teachers' results. Teachers score if their answers agree with the so-called aggregate expert.

Results

Expert-Rating

Results of the Model Items

For creating a point of reference, the median of classified data for each answer of the model items was calculated with IBM® SPSS® Statistic. The median of classified data is used as the so-called aggregate expert of the model items. After this, the four answers of each item were ranked (see Tab. 4 and 5). The rank of an answer describes the position in the order that was done after the ranking. The order was the basis for creating quasi-ranked pairs (Thillmann, 2008, Walpuski et al., 2012) (Tab. 4 and 5). For the fifteen model items fifteen orders and 84 quasi-ranked pairs (relations) were created (booklet A: 36 relations and one item is a true/false item; booklet B: 48 relations).

Scores for the experts were distributed based on the relations. An expert or test-person will score if he or she estimates the same relation as the aggregate experts (Walpuski et al., 2012) (Tab. 5). If the difference of the median of classified data of two answers is smaller than .50, as shown in Tab. 4, the rank is not distinct. The expert or test person will score, if he or she estimates e.g. A1 lower or the same as A2 on the Likert-scale, but he or she will not score if A1 is rated higher than A2, A3, or A4 (see Tab. 6).

8 relations of booklet A and 8 relations of booklet B show a variance of zero that is why they are not included in the calculations (booklet A: $\alpha = .807$; $n_{\text{relations/one item}} = 29$; $n_{\text{experts}} = 9$; booklet B: $\alpha = .430$; $n_{\text{relations}} = 40$; $n_{\text{experts}} = 9$). Cronbach's alpha was calculated by using data, in which missing data were recoded into zero points as it is common for achievement tests.

Table 4. Order and relations of Item MMp 4.

Answer	MD_{cd}	Rank	Order	Relations
A1	2,29	1	A1 ≤ A2 < A3 ≤ A4	A1 ≤ A2;
A2	2,33	2		A1 < A3; A1 < A4
A3	5,33	3		A2 < A3; A2 < A4;
A4	5,17	4		A3 ≤ A4

MD_{cd} = Median of classified data. Notice that the quasi-ranked pairs A1≤ A2 and A3 ≤ A4 contain an equal to or less than sign. If teachers rank A1 equal to or less than A2 they get a point.

Table 5. Order and relations of Item MNM 28.

Answer	MD_{cd}	Rank	Order	Relations
A1	1,50	1	A1 < A4 < A3 < A2	A1 < A2;
A2	4,75	4		A1 < A3; A1 < A4;
A3	4,17	3		A2 > A3; A2 > A4;
A4	3,00	2		A3 > A4

MD_{cd} = Median of classified data.

Table 6. Examples of possible teachers' ranking and points of item MNM 28.

	A1	A2	A3	A4	Order	Points
Teacher 1	1	5	4	2	A1<A4<A3<A2	6
Teacher 2	2	4	3	1	A4<A1<A3<A2	5

The frequency distribution of the scores was calculated (see an example in Tab. 7). A relation is not included in the calculations later, e.g. MNM 28 relation 2_3 (Tab. 7), if less than 5 experts score in a relation. At the end booklet A contains 32 relations and booklet B 43 relations.

The two items which have no Likert-scale and had to be answered by a ranking were interpreted differently. The item in which the steps of a process have to be ranked has just one right answer based upon the theoretical background. The right answer gets one point. This item is part of booklet A which is why booklet A has 36 relations at the beginning. The point of reference for the other item which involves the structure of acquiring knowledge by using models refers to

Table 7. Frequency distribution of the relations of Item MMp 4 and MNM 28.

Item label and relation	*f* (1 Point)
Mmp 4 relation 1_2	5
Mmp 4 relation 1_3	9
Mmp 4 relation 1_4	8
Mmp 4 relation 2_3	6
Mmp 4 relation 2_4	8
Mmp 4 relation 3_4	7
MNM 28 relation 1_2	9
MNM 28 relation 1_3	9
MNM 28 relation 1_4	8
MNM 28 relation 2_3	4
MNM 28 relation 2_4	9
MNM 28 relation 3_4	6

FD = Frequency distribution, $N_{\text{experts}} = 9$. The frequency distribution shows the number of experts who got a point for these relations.

relations as described above. This item is contained in booklet B that is why booklet B has 48 relations.

Results of the Chemistry Language Items

The expert-rating was used for creating a point of reference for the 60 possible activities (15 items respectively four answer alternatives). For each possible activity, it was analysed if the aggregate expert agrees or refuses the possibility of action. The median of each answer alternative was calculated and used as the aggregated experts' opinion. On this account, the Likert-scale (1 to 6) was dichotomized (1 to 3 "experts' agreement", 4 to 6 "experts' rejection"). The expert-rating shows a good internal consistency of $\alpha = .81$.

Pilot Study

Results of the Model Items

32 relations and one item of booklet A are used for teacher's scoring. 8 relations of this booklet were deleted because of a negative separation effect. One relation had a variance of zero. The mean value score of booklet A is 15.38 ($N_{teachers}$ = 26; Min = 2, Max = 23, $s^2 = 19.52$; $SD = 4.42$). The reliability of booklet A is satisfying ($\alpha = .816$, $n_{\text{relations/one item}} = 23$; $N_{teachers}$ = 23 (3 teachers were deleted because of missing data)).

43 relations of booklet B are used for teacher's scoring. 11 relations of this booklet were deleted because of a negative separation effect. Booklet B has a satisfying Cronbachs alpha ($\alpha = .864$, $n_{\text{relations}} = 32$). These calculations are based upon 15 teachers. 8 teachers were excluded because of missing data. The mean value score of booklet B is 22.91 (n_{teachers} = 23; Min = 12; Max = 31; $s^2 = 27.68$; $SD = 5.26$).

The sum scores of both booklets are distributed normal which was tested by the Kolmogorov-Smirnov-Test.

Even though relations which have a negative separation effect were not included in the calculations, there are relations with a bad separation effect (booklet A: .086 to .601; booklet B: .000 to .607). Relations, which show a separation effect lower than .3 were revised by formulating the alternative answers much more clearly. This means if a relation which has a bad separation effect is A1 < A2, A1 and A2 were revised in a way that the differentiation between the answer alternatives is much more strength. A rating of these answer alternatives should be much more easier which means that one answer should be better than the other one.

Results of the Chemistry Language Items

Teachers' rated the answers on a Likert-scale (1 to 6), as aforementioned. Afterwards the Likert-scale was dichotomized (1 to 3 "agreement", 4 to 6 "rejection"). Based upon the mean values of the expert-rating an appraisement of teachers' rating was done. Teachers' scoring is distributed normal, which was tested by a Kolmogorov-Smirnov-test. However, neither booklet A ($\alpha = .511$) nor booklet B ($\alpha = .357$) have satisfactory reliabilities. The unsatisfactory reliabilities may be accounted for by the small number of items (booklet A: $n_{\text{item}} = 32$ (2 without any variance), booklet B: $n_{\text{item}} = 28$ (4 without variance) and a small number of testpersons (booklet A: $n_{\text{teacher}} = 26$, booklet B: $n_{\text{teacher}} = 23$) (Wirtz and Caspar, 2002).

In addition to the results of the pilot-study, the results of the expert-rating were taken into account to evaluate the chemistry language items. The items were evaluated by using different criteria originating in the data of the pilot study and the expert-rating. These criteria are the item-separation and the standard deviation of the experts' answers regarding the answer alternatives. Furthermore, items which have no variance in the pilot study were identified. Assuming that the selected items are not adequate for explaining variance or obtaining a majority for refusing or agreeing to an alternative answer, these items were radically revised. Moreover, qualitative and individual modification proposals of the experts were included in the revision. Statements and answers which were linguistically unclear have been rephrased and items with an unsatisfactory separation value were revised. The dialogs were not changed in the revision process.

In this manner 12 items with 48 answer alternatives were devised. These 12 items, 12 model items and 14 experiments items were arranged in the new booklet FEMo.

Discussion and Outlook

All in all, pilot study's results are encouraging. The systematic approach for measuring chemistry teachers' PCK in terms of dealing with models and chemistry language seems to be helpful and lead to a reliable test instrument that is adequate for large samples. The recent results are used in a constructive manner in order to improve items and answer alternatives. One substantive critical issue could be the small test person sample of each booklet ($n_A = 26$, $n_B = 23$) of the pilot study, which was reduced by missing data (Wirtz and Caspar, 2002). Furthermore, the expert sample is very small and consists of nine professors and one teacher trainer, only. A larger expert sample would be better not only for an explicit agreement or rejection of the answer alternatives, but also for distinct relations.

Another problem is the high number of excluded relations of the model items. It reduces scoring for the items in a substantive way. For example only for three relations of item MMp 4 a scoring could be done. This reduces total scoring from 6 to 3 points for this item. In addition, scoring for every item is different. Some items just have a maximum number of 2 points whereas others have a maximum number of points of 6. Nevertheless, analysing model items by using relations is a good alternative to a right and false analysis, because acting in class is influenced by different aspects that is why there is no right or false behaviour in class.

Validity of the FEMo has not been analysed, so far. Vogelsang and Reinhold (2013) emphasize the importance of analysing professional knowledge tests for their ability to measure the knowledge of behaviour in class (validity of action). Expert-ratings and think-aloud interviews can be used for analysing the validity of action, even though these test processes are indirect (Jüttner and Neuhaus, 2013;

Vogelsang and Reinhold, 2013). Jüttner and Neuhaus (2013) could analyse the content validity of their test successfully by using think-aloud interviews. Krauss, Baumert, and Blum (2008), and Dollny (2011) used contrast group for validation. According to Dollny (2011) and Jüttner and Neuhaus (2013) it is planned to analyse the validity of the FEMo by interviewing and testing science teachers (biology and physics) and non-science teachers (English, German and/or French) as a contrast group. The teachers will answer the complete FEMo booklet, containing model items, experiment items, and chemistry language items. Biology and physic teachers use models and experiments often in a different way to chemistry teachers. That is why it is interesting to analyse if items focusing on these two facets could differentiate between chemistry specific PCK of using models and experiments in class and PCK of biology and physic teachers. Teachers' results will be compared to results of chemistry teachers of the main study.

Before starting the validation a new expert-rating will be conducted, because the revision of items could have changed the agreement or rejection of alternative answers. For this purpose it is planned to ask professors of chemistry education to serve as experts, again.

All in all first results of the main study are expected for June 2014 and we are confident of being able to present a reliable and valid PCK test instrument (FEMo) at the end of the study.

References

Baumert J. and Kunter M., Stichwort: Professionelle Handlungskompetenz (Keyword: Professional Competence), *Zeitschrift für Erziehungswissenschaften*, **9**(4), 469-520, 2006. Retrieved from: http://www.schulentwicklung.bayern.de/unterfranken/userfiles/SETag2010/Baumert.pdf

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. and Tsai, Y.-M., Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress, *American Educational Research Journal*, **47**(1), 133-180, 2010. DOI: 10.3102/0002831209345157

Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach: Sprachlichkeit und fachliches Lernen (Subject-Didactics' Research: Vol. 3 Language in Subject: Linguisticality und Subject-Specific Learning)*, Münster [u.a.]: Waxmann, 2013.

Buhlmann, R., and Fearns, A., *Handbuch des Fachsprachenunterrichts: Unter besonderer Berücksichtigung naturwissenschaftlich-technischer Fachsprachen (Handbook of Technical Language Teaching: A Peculiar Consideration of Science-and-Technics-Related Technical Languages)* (6th ed.), *Narr-Studienbücher*. Tübingen: Narr, 2000.

Busch, H. and Ralle, B., Diagnostik und Förderung fachsprachlicher Kompetenzen im Chemieunterricht (Diagnosis and Fostering of Technical Language Competencies in Chemistry Classes). In: Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 277–295). Münster [u.a.]: Waxmann, 2013.

Bußmann, H. (ed.), *Lexikon der Sprachwissenschaft (Dictionary of Linguistics)* (3rd ed.). Stuttgart: Kröner, 2002.

Dollny, S., *Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften (Development and Evaluation of a Testinstrument for Measuring Chemistry Teachers Subject Specific Professional Knowledge)*. Berlin: Logos, 2011.

Eunson, B., *Communicating in the 21st century* (3. ed). Milton: Wiley, 2012.

Fluck, H.-R. (ed.), *Uni-Taschenbücher: 483: Germanistik, Linguistik. Fachsprachen: Einführung und Bibliographie (Uni-Paperbacks: 484: German studies, Linguistics, Technical Language: Introduction and Bibliography)* (5th ed.). Tübingen [u.a.]: Francke, 1996.

Garritz, A., PCK for dummies, *Educación Química*, **24**(2), 462-465, 2013.

Gess-Newsome, J., The PCK Summit Consensus Model and Definition of Pedagogical Content Knowledge. In: The Symposium "Reports from the pedagogical Content Knowledge (PCK) Summit", ESERA Conference 2013, September, 2013.

Gilbert, J. K., Models and Modelling: Routes To More Authentic Science Education, *International Journal of Science and Mathematics Education*, **2**(2), 115-130, 2004.

Grosslight, L., Unger, C. and Jay, E., Understanding Models and their Use in Science: Conceptions of Middle and High School Students and Experts, *Journal of Research in Science Teaching*, **28** (9), 799-822, 1991.

Grucza, S., *Fachsprachenlinguistik (*Linguistics of Technical Languages*)*. Frankfurt am Main: Lang, 2012.

Harrison, A. G., and Treagust, D. F., A typology of school science models, *International Journal of Science Education*, **22**(9), 1011-1026, 2000.

Henze, I., Van Driel, J., and Verloop, N., Science Teachers' Knowledge about Teaching Models and Modelling in the Context of a New Syllabus on Public Understanding of Science, *Research of Science Education*, 37, 99-122, 2007a. DOI: 10.1007/s11165-006-9017-6

Henze, I., Van Driel, J., and Verloop, N., The Change of Science Teachers' Personal Knowledge about Teaching Models and Modelling in the Context of Science Reform, *International Journal of Science Education,* **29**(15), 1819-1846, 2007b. DOI: 10.1080/09500690601052628

Justi, R. S. and Gilbert, J. K., Modelling teachers' view on the nature of modelling, and implications for the education of modellers, *International Journal of Science Education*, **24**(4), 369-387, 2002a.
DOI: 10.1080/09500690110110142

Justi, R. S. and Gilbert, J. K., Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science, *International Journal of Science Education*, **24**(12), 1273-1292, 2002b.
DOI: 10.1080/09500690210163198

Justi, R. S. and Gilbert, J. K., Teachers‘ views on the nature of models, *International Journal of Science Education*, **25**(11), 1369-1286, 2003. DOI: 10.1080/0950069032000070324

Justi, R. and van Driel, J., The development of science teachers’ knowledge on models and modelling: promoting, characterizing, and understanding the process, *International Journal of Science Education*, **27**(5), 549-573, 2005. DOI: 10.1080/0950069042000323773

Jüttner, M., Boone, W., Park, S., and Neuhaus, B. J., Development and use of a test instrument to measure biology teachers‘ content knowledge (CK) and pedagogical content knowledge (PCK), *Educational Assessment, Evaluation and Accountability*, **25**(1), 45-67, 2013. DOI: 10.1007/s11092-013-9157-y

Jüttner, M., *Entwicklung, Evaluation und Validierung eines Fachwissenstests and fachdidaktischen Wissenstests für die Erfassung des Professionswissens von Biologielehrkräften (Development, Evaluation and validation of a content knowledge test and a pedagogical content knowledge test measuring biology teachers‘ knowledge)*, Doctoral Thesis, Biology Education. LMU Munich, Germany, 2013.

Jüttner, M. and Neuhaus, B. J., Validation of a Paper-and-Pencil Test Instrument Measuring Biology Teachers‘ Pedagogical Content Knowledge by Using Think-Aloud Interviews, *Journal of Education and Training Studies*, **1**(2), 113-125, 2013. DOI: 10.1114/jets.vli2.126

Kessel, K., and Reimann, S., *Basiswissen deutsche Gegenwartssprache (Fundamental Knowledge of Contemporary German Language)* (2nd ed.), *UTB: 2704: Sprachwissenschaft*, Tübingen [u.a.]: Francke, 2008.

Kirschner, S., *Modellierung und Analyse des Professionswissens von Physiklehrkräften (Modelling and Analysis of physic teachers‘ professional knowledge)*. Berlin: Logos, 2013.

KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (eds.) *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004 (Enactment of the conference of the secretaries of education – education standards in the subject chemistry for the middle graduation)*. München, Neuwied: Luchterhand, 2005.

Koch, P. et Oesterreicher, W., Langage de la proximité - langage de la distance. Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte: Langage de la proximité - langage de la distance (Language of Proximity - Language of Distance. Oral and Writing in in the Area of Conflict of Language Theory and Language History), L’oralité et la scripturalité entre la théorie linguistique et l’histoire de la langue, *Romanistisches Jahrbuch*, 36, 15–43, 1985.

Koch, P. and Oesterreicher, W., Schriftlichkeit und kommunikative Distanz (Writing and Communicative Distance), *Zeitschrift für Germanistische Linguistik*, **35**(3), 2007. DOI:10.1515/ZGL.2007.024

Krauss, S., Baumert, J. and Blum, W., Secondary mathematics teachers‘ pedagogical content knowledge and content knowledge: validation of the COACTIV constructs, *ZDM The International Journal on Mathematics Education*, 40, 873-892, 2008. DOI: 10.1007/s11858-008-0141-9

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M. and Jordan, A., Pedagogical Content Knowledge of Secondary Mathematics Teachers, *Journal of Educational Psychology*, **100**(3), 716-725, 2008. DOI: 10.1037/0022-0663.100.3.716

Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. and Jordan, A., Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie (Analysis of German mathematic teachers‘ Professional Knowledge in the framework of the COACTIV Study), *Journal für Mathematik-Didaktik*, **29**(3/4), 223-258, 2008.

Kulgemeyer, C. and Schecker, H., Schülerinnen und Schüler erklären Physik – Modellierung, Diagnostik und Förderung von Kommunikationskompetenz im Physikunterricht (Students Explain Physics – Modelling, Diagnosis and Fostering of Communicational Competencies in Physics Classes). In: M. Becker-Mrotzek (Ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 225–241). Münster [u.a.]: Waxmann, 2013.

Kunter, M., Kleickmann, T., Klusmann, U. and Richter, D., Die Entwicklung professioneller Kompetenz von Lehrkräften (Development of teachers‘ professional competence). In: Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. and Neubrand, M. (ed.): *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV (Professional competence of teachers. Results of the research program COACTIV)* (pp. 55–68, Münster: Waxmann, 2011.

Linneweber-Lammerskitten, H., Sprachkompetenz als integrierter Bestandteil der mathematical literacy? (Language Competence as Integrated Element of Mathematical Literacy?). In: Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 151–167). Münster [u.a.]: Waxmann, 2013.

Maia, P. F. and Justi, R., Learning of Chemical Equilibrium through Modelling-based Teaching, *International Journal of Science Education*, **31**(5), 603-630, 2009. DOI: 10.1080/09500690802538045

Merzyn, G., Sprache und Chemie lernen (Language and Chemistry Learning), *Naturwissenschaften im Unterricht*, **19**(106/107), 94–97, 2008

Mikelskis-Seifert, S., Lernen über Modelle. Am Beispiel der Teilchenstruktur der Materie (Learning about models. Using the example of particles), *Praxis der Naturwissenschaften Chemie in der Schule*, **7**(58), 15-21, 2009.

National Research Council, *National Science Education Standards*, Washington DC: National Academy Press, 1996.

Norris, S. P. and Phillips, L. M., How literacy in its fundamental sense is central to scientific literacy, *Science Edu-*

cation, **87**(2), 224–240, 2003. DOI:10.1002/sce.10066
Özcan, N., Zum Einfluss der Fachsprache auf die Leistung im Fach Chemie: Eine Förderstudie zur Fachsprache im Chemieunterricht. Studien zum Physik- und Chemielernen: Vol. 144 (To the Influence of Chemistry Language on the Performance in Chemistry Classes: A Research Study to Foster Chemistry Language in Chemistry Classes). Berlin: Logos Berlin, 2013.
Parchmann, I. and Bernholt, S., In, mit und über Chemie kommunizieren – Chancen und Herausforderungen von Kommunikationsprozessen im Chemieunterricht (Communicate in, with and about Chemistry. Chances and Challenges of Communicational Processes in Chemistry Classes). In: Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 241–255). Münster [u.a.]: Waxmann, 2013.
Pollender, T. (unpublished results). Doctoral dissertation in preparation. Chemistry Education. University of Duisburg-Essen.
Riese, J. and Reinhold, P., Entwicklung und Validierung eines Instruments zur Messung professioneller Handlungskompetenz bei Physiklehrern (Development and Validation of a Test-instrument for Measuring Physics Teachers‘ Competence), *Lehrerbildung auf dem Prüfstand*, **1**(2), 625-640, 2008.
Riese, J. and Reinhold, P., Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen (The Professional Competence of Prospective Physics Teachers in Different Forms of Qualification), *Zeitschrift für Erziehungswissenschaft*, **15**(1), 111–143, 2012. DOI:10.1007/s11618-012-0259-y
Rincke, K., *Sprachentwicklung und Fachlernen im Mechanikunterricht: Sprache und Kommunikation bei der Einführung in den Kraftbegriff. Studien zum Physik- und Chemielernen: Vol. 66 (Development of Language and Subject-Related Learning in Mechanic-Classes: Language and Communication with the Introduction of the Term Force)*. Berlin: Logos-Verl, 2007.
Roelcke, T., *Fachsprachen (Technical Languages)* (2., durchges. Aufl), *Grundlagen der Germanistik: Vol. 37*, Berlin: Schmidt, 2005.
Saari, H. and Viiri, J., A research-based teaching sequence for teaching the concept of modelling to seventh-grade students, *International Journal of Science Education*, **25**(11), 1333-1352, 2003.
DOI: 10.1080/0950069032000052081
Schmelzing, S., Wüsten, S., Sandmann, A. and Neuhaus, B., Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung (Pedagogical content knowledge and reflection in frame of biology teacher education), *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 189-207, 2010. Retrieved from: http://www.ipn.uni-kiel.de/zfdn/pdf/16_Schmelzing.pdf
Schmölzer-Eibinger, S., Sprache als Medium des Lernens im Fach (Language as Medium of Learning in Subjects). In Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 25–41). Münster [u.a.]: Waxmann, 2013.
Shulman, L., Those Who Understand: Knowledge Growth in Teaching, *Educational Researcher*, **15**(2), 4-14, 1986.
Shulman, L., Knowledge and Teaching: Foundations of the New Reform, *Harvard Educational Review,* **57**(1), 1-22, 1987.
Sumfleth, E., Kobow, I., Tunali, N., and Walpuski, M., Fachkommunikation im Chemieunterricht (Subject-Related Communication in Chemistry Classes). In: Becker-Mrotzek, M. (Ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 255–277, Münster [u.a.]: Waxmann, 2013.
Sumfleth, E. and Pitton, A., Sprachliche Kommunikation im Chemieunterricht: Schülervorstellungen und ihre Bedeutung im Unterrichtsalltag (Linguistic Communication in Chemistry Classes: Students' Conceptions and their Meaning for Everyday Teaching), *Zeitschrift für Didaktik der Naturwissenschaften*, **4**(2), 4–20, 1998. Retrieved from ftp://ftp.rz.uni-kiel.de/pub/ipn/zfdn/1998/Heft2/S.4-20_Sumfleth_Pitton_98_H2.pdf
Tepner, O., Backes, A. and Sumfleth, E., Test zum fachdidaktisch-experimentellen Wissen von Chemielehrkräften (Pedagogical Content and Experimental Knowledge Test for Chemistry Teachers), unpublished manuscript, 2012.
Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Thillmann, H. and Wirth, J., Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften (Item Development Model for Assessing Professional Knowledge of Science Teacher), *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7-28, 2012. Retrieved from: http://www.ipn.uni-kiel.de/zfdn/pdf/18_Tepner.pdf
Tepner, O. and Dollny, S., Measuring Chemistry Teachers' Content Knowledge: Is It Correlated to Pedagogical Content Knowledge. In: C. Bruguiere A. Tiberghien and P. Clément (Eds.), *Topics and Trends in Current Science Education*, 9th ESERA Conference Selected Contributions (pp. 243–254). 2014.
Thillmann, H. and Wirth, J., unpublished manuscript. Retrieved from: https://www.uni-due.de/prowin/phase1-phy.shtml
Thillmann, H., Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung (Self-regulating learning through experimentation: From the registration to the advancement), University of Duisburg-Essen, 2014. Retrieved from http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-18970.xml
Troeger, H., Struebe, M., Sumfleth, E. and Tepner, O., Professional Knowledge of Chemistry Teachers Video Analysis of Chemistry Lessons. In C. P. Constantinou, N. Papadouris and A. Hadjigeorgiou (eds.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For*

Evidence-based Teaching and Coherence in Learning. Part 3 (co-ed. Tiberghien, A. and Kyza, E. A.) (pp. 83-88). Nicosia, Cyprus: European Science Education Research Association, 2014. Retrieved from http://www.esera.org/media/esera2013/Holger_Troeger_16December20131.pdf

Van Driel, J. H., de Jong, O. and Verloop N., The Development of Preservice chemistry Teachers' Pedagogical Content Knowledge, *Science Education*, **86**(4), 572-590, 2002. DOI: 10.1002/sce.10010

Vogelsang, C. and Reinhold, P., Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften (Concerning the validity of assessments of science teachers' professional knowledge), *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 103-128, 2013. Retrieved from: http://www.ipn.uni-kiel.de/zfdn/pdf/19_Vogelsang.pdf

Vollmer, H. J. and Thürmann, E., Sprachbildung und Bildungssprache als Aufgabe aller Fächer der Regelschule (Language Formation and Language of Education as Responsibility of all Subjects of Regular Schools). In: Becker-Mrotzek, M. (ed.), *Fachdidaktische Forschungen: Vol. 3. Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 41–59). Münster [u.a.]: Waxmann, 2013.

Walpuski, M., Tepner, O., Sumfleth, E., Dollny, S., Hostenbach, J. and Pollender, T., Multiple perspectives on students' scientific communication and reasoning in chemistry education, *Acta Didactica Norge*, **6**(1) Art. 11, 2012. Retrieved from http://adno.no/index.php/adno/article/view/206/250

Wellington, J., and Osborne, J., *Language and literacy in science education*, Buckingham Philadelphia: Open University Press, 2001.

Wirtz, M. A. and Caspar, F., *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen (Rating-agreement and rater-reliability: Methods for assessing and improving reliability of ratings via category systems and rating-scales)*, Göttingen [u.a.]: Hogrefe, Verl. für Psychologie, 2002.

Witner, S. and Tepner, O., Entwicklung geschlossener Testaufgaben zur Erhebung des fachdidaktischen Wissens von Chemielehrkräften (Development of Multiple-Choice Items for Measuring Teachers' Pedagogical Content Knowledge), *Chimica et ceterae artes rerum naturae didacticae*, **37**(104), 113-137, 2011.

Yore, L. D. and Treagust, D. F., Current Realities and Future Possibilities: Language and science literacy — empowering research and informing instruction, *International Journal of Science Education*, **28**(2-3), 291–314, 2006. doi:10.1080/09500690500336973