Maik Walpuski, professor
Oliver Tepner
Elke Sumfleth, professor
Sabrina Dollny
Julia Hostenbach
Tobias Pollender
University Duisburg-Essen

VISIONS
Conference
2011:
Teaching

Multiple perspectives on students' scientific communication & reasoning in chemistry education

Abstract

Both students and teachers need different competences for scientific reasoning. Apart from the required content knowledge and the ability of using it adequately, both groups need elaborated knowledge of strategies for decision-making and argumentation. These competences concerning decision-making are highly dependent on how frequently students are given the chance to argue in science classes. This article pools the results of three different research projects in chemistry education which focus on these three aspects: Students' competences, the classroom situation with regard to communication and reasoning and teachers' competences. Students' skills needed for discussing chemistry content and for decision-making in scientific contexts are analyzed first. Following this, the opportunities for improving these skills during science lessons are researched and related to results from a study of teachers' pedagogical content knowledge (PCK) in this domain. The analysis of students' and teachers' communication skills is conducted in two different ways, paper-pencil tests and video analysis.

Paper-pencil tests are used to assess students' performance in reasoning. The test items deal with chemistry-specific situations, including opportunities for decision-making, which are typical of socio-scientific issues (study 1). The study shows that students' decision-making skills are poor when the topics deal with scientific contexts, but that students perform better when dealing with everyday-life contexts. One reason might be the lack of reasoning in chemistry lessons, as shown by a video study conducted in chemistry classes. Students' and teachers' in-class behavior and communication patterns are analyzed with regard to students' and teachers' contributions (study 2). The amount and the quality of students' and teachers' statements as well as the interactions, especially teachers' reactions to students' statements are investigated. Teachers' way of negotiating and guiding classroom discussions should depend on their PCK in this specific field. Due to a lack of studies on correlations between teachers' PCK and classroom activities, the third study focuses on the development of adequate tests on PCK. Tests on content knowledge (CK) are developed in addition to prove the expected correlations between PCK and CK (Baumert & Kunter, 2006).

Introduction

The concept of scientific literacy emphasizes that science plays a key role in everyday life. Scientific knowledge leads to a deeper and more sophisticated understanding and describing of everyday-life situations and can be seen as an important part of general education (Bybee, 2002; Bybee, Fensham, & Laurie, 2009; DeBoer, 2000; KMK, 2005; Laugksch, 2000). Science educators worldwide have pointed out that scientific reasoning is an important issue of science education (AAAS Project 2001, 2001; AAAS Project 2061, 2007; National Research Council, 1996; Sadler & Zeidler, 2005a). Therefore, the acquisition of content knowledge in and of itself cannot be the only aim of schooling. Students need skills in addition to content knowledge to make evidence-based decisions in science contexts, e.g. knowing and applying strategies.

Matching these requirements, the National Educational Standards in Germany focus on four different areas of competence: *Content knowledge, acquirement of knowledge, evaluation and judgement*, and *communication*. The two areas of competence *evaluation and judgement* and *communication* especially affect socio-scientific issues (SSI), which are assumed to form a link between informal reasoning skills and content knowledge (Sadler & Zeidler, 2005b).

The studies presented in this paper investigate students' skills in decision-making in scientific contexts as well as teachers' skills in fostering these skills in students. The paper delves firstly into whether decision-making is a domain-specific competence, investigating what other factors of knowledge and behavior influence students' scientific decision-making. For this purpose, students' skills on evaluation strategies, content knowledge, cognitive abilities and their attitudes are related to their performance in decision-making on socio-scientific issues. Secondly, classroom discussions are examined via video analyses to find out how teachers react to students' arguments, especially in terms of dealing with students' pre-conceptions, and how they foster students' argumentation. At the same time, students' opportunities for developing reasoning and decision-making competences are examined. An important aspect is whether teachers encourage their students to argue their statements.

The data from these video analyses will, thirdly, be compared to the teachers' PCK which is investigated in the last study. It could be assumed that a deficient knowledge of students' preconceptions (as a main facet of PCK) may result in teacher-dominated communication processes hiding uncertainty in open discussions. Additionally, encouraging students' explanations and supporting decision-making in complex situations are not fostered (Van Dijk & Kattmann, 2010). In order to gather more information about this lack of knowledge, chemistry-specific PCK and CK tests referring to knowledge of students' preconceptions have been developed.

Generally, there are two ways of exploring students' abilities in scientific reasoning: Process analyses (e.g., by video analyses or interviews) can be used to investigate students' behavior in detail, while summative assessments can be used to identify and measure the competences necessary for scientific reasoning and decision making.

Study 1: Assessing students' evaluation and judgement competences

Background

The study (Göbel & Walpuski, 2010; Hostenbach, 2011) was carried out in the context of the Evaluation of German National Educational Standards (KMK, 2005). The German National Educational Standards are based on the internationally accepted concept of scientific literacy (Neumann, Fischer, & Kauertz, 2010) and on Weinert's (2001) definition of competence. As they provide the framework of the study, they will be briefly introduced in the following paragraph.

The PISA findings that revealed shortcomings in German science education instigated a broad discussion on possible changes in science education. At that time, the German education system was characterized by an input orientation with curricula describing in detail which content should be taught; National Educational Standards (NES) and national assessments for describing and measuring outcome had not been issues in educational administration. The result of the discussion on PISA findings was a structural reform of the educational system, which included the implementation of NES and national assessments (Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010; Neumann et al., 2010; Walpuski, Ropohl, & Sumfleth, 2011). Based on the concept of scientific literacy, the NES for science education define four different areas of competence which contribute to gaining scientific literacy. As stated in the introduction, the four areas of competence are *content knowledge*, *acquisition of knowledge*, *communication* and *evaluation and judgment*. The standards are defined for the natural sciences (biology, chemistry, and physics) separately and have been in effect in Germany since the school year 2005/2006. The NES are defined for the graduation from middle schools after grade 9 or 10 (depending on the school type).

Reasoning skills play a role in two areas of competence: *Communication* and *evaluation and judgment*. *Evaluation and judgement competence* in chemistry is defined as "the ability to detect and evaluate chemical topics in different contexts" (KMK, 2005) and focuses on informal reasoning skills, such as evaluating different options of behavior, finding reasons for and against different options and reflecting on possible consequences. Students should be able to justifiably and systematically choose between different options of behavior in complex problem-situations (Eggert & Bögeholz, 2010). The

communication competence focuses on the process of argumentation (e.g., how to support a statement with evidence), among others.

It can be assumed that there are different abilities that are necessary for making decisions on socio-scientific issues. Kortland (1996) describes decision-making in a normative model as a circular process, which consists of the identification of a problem, the development of criteria, the generation and evaluation of alternatives and the final solution. This is more or less identical to the description of evaluation and judgement competence in the German NES.

Currently, students' decision-making processes concerning SSI (socio-scientific issues) in different scientific contexts are not well evaluated. SSI are understood in the way they are defined by Sadler (2004). SSI are complex, open-end problems and they contain more than one correct solution. SSI always represent a situation where students have to decide, in order to improve the ability of decision-making and to improve the selection of one option (Uskola, Maguregi, & Jiménez-Aleixandre, 2010). SSI comprise problems, topics and dilemmas, which are related to science and scientific knowledge. In addition, scientific practice is useful for negotiating SSI. During handling SSI, not only the solution is important, but also the process of decision-making is fundamental for the ability of dealing with SSI. If the process of decision-making is made explicit by e. g. pointing out the criteria used, reasoning skills become necessary and obvious.

Our assumption is that students need the following personal, inter-disciplinary, and subject-specific abilities for evaluating SSI:

1. To know evaluation strategies (interdisciplinary)
2. To be able to apply evaluation strategies (interdisciplinary/everyday-life)
3. To have domain-specific content knowledge (subject-specific)
4. To be able to apply knowledge to a given situation (subject-specific).

The evaluation strategies are assumed to be necessary as a precondition for taking reasonable decisions. Students should know that a justified decision needs to meet criteria which have to be found in a first step. Additionally, students have to know that there are different ways to deal with the criteria. They can e.g. be compared and ranked (trade-off) or attributes can be defined as indispensable (cut-off) so that they have to be reached and a trade-off is impossible. In addition to knowing these strategies, students have to be able to apply them – in the easiest case – on everyday-life problems (2). The evaluation of criteria in subject-specific contexts always depends on the domain-specific knowledge, which is required e.g. to decide which criterion is more or less important or which argument is incorrect from a scientific point of view (3). Similar to the application of strategies, having the content knowledge is not enough; students do have to be able to apply it to the concrete situation.

Additionally, students' reasoning and decision-making processes may be influenced by the following personal aspects:

1. Cognitive abilities.
2. Individual attitudes.
3. Social desirability of their answer.

The first factor is chosen because cognitive abilities are in general a strong predictor for achievement. Since for ranking criteria not only knowledge about the topic plays an important role but individual attitudes may play an additional role, they are assessed, too. Due to the fact that we assess the evaluation and judgement on SSI students may show a social desirable reaction e.g. regarding environmental problems. For this reason, the tendency to give social desirable answers is assessed with an existing questionnaire.

The aims of this study are a) to investigate whether the evaluation and judgement competence is domain-specific and b) to find out to which degree the different assumed factors explained above influence students' scientific decision-making.

Methods

Due to the aim of the study, a quantitative empirical design was used. The core of the study was a newly developed test assessing students' reasoning and decision-making in science contexts. The test is a paper-pencil test. In the items, different aspects of competence are represented systematically. The items are assigned to of the following sub-competences:

1. Evaluating criteria for decision-making in scientific contexts (scientific reasoning)
2. Evaluation of different options of behavior
3. Reflection of the results.

The items address SSI problems either from a chemistry or a biology perspective in order to find out if there are any subject-specific differences. The test items are multiple choice, short answer or open answer questions and are coded either full credit or no credit in order to keep the different item formats comparable. Factors influencing the items difficulty (e.g. the complexity of the problem) are systematically varied. The items include a description of the situation as well as background information on relevant chemical or biological data that are important for making a decision. Similar items – concerning everyday-life topics – were used to measure students' general evaluation and judgement competence and to compare it to the items from the scientific contexts. An example for an item from a chemistry perspective is e.g. that students have to decide if they prefer a car running on bio-ethanol compared to a car running on regular fuel.

The emissions and energy for the production and the use of the fuel have to be taken into account. Another example addressing the problem of selecting chemicals for a given purpose can be found in Figure 1.

Item stem: Cornelia is injured and now she wants to create a cooling pack. For this, she needs salt and water. The dissolution of salt in water is a process, which is endothermic or exothermic. To choose a salt type, Cornelia looks at the salts' dissolution energies and prices. The cooling pack should be as cold and cheap as possible.

The dissolution energy can be:

- negative: the solution gets warm, the dissolution is exothermic
- zero: the solutions' temperature doesn't change
- positive: the solution gets cold, the dissolution is endothermic

The temperature of 50 mL water increases by about 3 °C when dissolving 5 g sodium iodide.

Salt	Dissolution Energy kJ/g	Price in €/100g
Calcium sulphate	- 0,13	3,63
Ammonium chloride	+0,27	2,73
Sodium chloride	+0,06	1,97
Sodium iodide	-0,05	19,30
Calcium fluoride	+0,23	0,79

Item: Which salt does Cornelia choose to create the cooling pack? Give a reason for your decision.

Figure 1: Example for a decision making item

To reach the full credit, the students have to evaluate two different criteria. The dissolution energy has to be positive and the price has to be low. The salt with the highest dissolution energy is relatively expensive compared to the one with the second highest dissolution energy. Since the task was to find a way to build the cold pack as cheaply as possible, the students have to select the calcium fluoride and to explain the trade-off for full credit.

In a comparable everyday-life situation students would have e.g. to decide which holiday to book – using criteria like price, expected weather and accommodation. As mentioned above, formal aspects which are known to cause difficulty, such as the complexity of the tasks (Kauertz & Fischer, 2006), were kept constant between the test items of different subjects. Additionally, tests concerning the different other aspects assumed to influence students' decision-making (content knowledge, cognitive abilities etc.) were administered to the students. During the process of decision-making criteria have to be identified, evaluation strategies have to be used, and finally a decision has to be made. Accordingly, the content knowledge and the application of evaluation strategies are important to evaluate and judge precisely. Different research projects show that high content knowledge influences the quality of argumentation (von

Aufschnaiter, Erduran, Osborne, & Simon, 2008) and creates a relation between content knowledge and the negotiation of socio-scientific issues (Sharan, 1980). To measure students' content knowledge, items from the same context as the decision-making tasks were added to the test booklets. The items either asked for factual knowledge or for knowledge to be applied to something. The cognitive ability test (Heller & Perleth, 2000) was used to control the influence of intelligence. To check whether the students make decisions because of the answers' social desirability, the balanced inventory of desirable responding was administered (Paulhus, 1998). A PISA questionnaire on individual attitudes concerning environmental aspects was used (Frey et al., 2009) to find out whether individual attitudes affect the decisions made. An overview of all these influencing factors is given in Figure 2.

Figure 2. *Factors influencing decision making*

The sample consisted of 750 students from the 9th and 10th grade of German upper secondary schools (Gymnasium). Different schools in North Rhine-Westfalia were randomly asked to participate in the study and the sample consisted of schools which responded to our request. For this reason, the selection of schools may be biased by interest but this is not problematic regarding the research questions. The sample always included complete classes from the participating schools.

To achieve a high validity, more items than one student could work on in a reasonable time were developed. As a result, a balanced incomplete block design for the items on decision-making and on content knowledge was used. This means that the items were combined to blocks and the blocks again to booklets. The different characteristics of the items were equally distributed over the blocks. Hence, approximately 140 students worked on each item. Resulting from the incomplete block design, data were analyzed based on probabilistic test theory and the Rasch model using the software Acer Conquest. The Rasch model is based on the assumption that, generally, students with a higher ability in the observed latent construct (e. g. evaluation and judgement competence) should score higher in a test than students with a lower ability in that construct. As a consequence, more difficult items are assumed to be answered correctly less often than easier ones. The results for different items are all transferred to the same scale, which is the main advantage of the Rasch model. As a result, measures for students' abilities (estimated person parameters) derived from different booklets (of varying difficulty) can be directly compared. To verify quality, the fit-parameters of the items and the item difficulty were investigated. All items appeared to be useful and of satisfying quality.

To find out if decision-making is a subject-specific competence, an uni-dimensional Rasch analysis was compared to a three-dimensional model, in which the items from chemistry, biology and everyday life contexts were scaled on different dimensions. Additionally, regressions and correlations were calculated to analyze how the different latent constructs are related.

Results

The first aim of the study was to investigate whether the evaluation and judgement competence is domain-specific. The comparison of the uni-dimensional Rasch model, which assumes the evaluation and judgement competence to be independent of the subject, with a three-dimensional Rasch model, which differentiates between biology, chemistry and everyday-life, shows that the latter has a better fit. This was done by using the AIC and the BIC (see Table 1). The Akaike information criterion (AIC) and the Bayesian information criteria (BIC) indicate the relative goodness of a statistical model's fit. Both AIC and BIC provide well-founded approaches to model comparison, but since both criteria can fail under certain conditions, Kuha (2004) recommends using the two criteria together to enhance the robustness of the choice. Given a set of possible models for the data, the preferred model is the one with the smaller AIC / BIC value. Evidence for a model is considered very strong if $BIC > 150$, strong if $BIC > 12$ and positive if $BIC > 3$ (Posada & Buckley, 2004).

Additionally, the decision is supported by the lower deviance for the three-dimensional model. In this case all three measures (AIC / BIC / Deviance) are lower for the three-dimensional model, which is therefore preferred. This means

that making decisions in biology, chemistry and everyday-life contexts requires different competences.

Table 1. *Model comparison for the evaluation and judgement competence*

	Uni-dimensional	Three-dimensional
<i>N</i>	777	777
Deviance	16110.96	16015.62
AIC	16306.96	16221.62
BIC	16677.48	16611.05

The latent correlations between the different latent constructs (dimensions) are still high, however, which shows that though the competences can be separated, they are still related to each other. The correlations are especially high between chemistry and biology, while the correlations of both with evaluation strategies in everyday-life are lower (see Table 2).

Table 2. *Latent correlations between the dimensions of the three-dimensional model*

	Everyday-life	Chemistry	Biology
Everyday-life	1		
Chemistry	0.547	1	
Biology	0.672	0.885	1

A comparison of the estimated person parameters shows that it is more difficult for the students to evaluate scientific situations than everyday-life situations. For this purpose, the mean of item parameters was set to zero for all three dimensions of the three-dimensional Rasch model. In a following step, the resulting person parameters were compared (Table 3). A person parameter of zero would mean that the student is able to solve 50 % of the test items on that dimension. A lower value describes a lower probability while a higher value describes a higher probability.

Table 3. *Estimated person parameters based on the three-dimensional model*

	Mean (<i>M</i>)	Standard Deviation (<i>SD</i>)
Everyday-life	2.04	1.50
Chemistry	.21	1.34
Biology	.51	1.68

The second aim of the study was to find out to which degree different factors influence students' scientific decision-making. Linear regressions show that most of the variance for the evaluation in chemistry can be explained by the ability to use evaluation strategies in everyday-life situations ($\beta = .21, p \leq .001$) followed by the ability to use content knowledge ($\beta = .19, p \leq .001$) and cognitive abilities ($\beta = .12, p \leq .05$).

In summary, it can be concluded that for scientific reasoning and decision-making, specific abilities are required. The ability to use evaluation strategies in everyday-life situations is one predictor, but content knowledge is needed in addition. Nevertheless the combination of both, decision-making competences in everyday-life and content knowledge is not sufficient for chemistry content-related decision-making. Therefore it is assumed that reasoning and decision-making in science contents have to be taught and practiced in chemistry classes.

Study 2: Analyzing in-class behavior and communication patterns

Insufficient student competences in evaluation and judgement may be caused by insufficient reasoning strategies and fewer opportunities for arguing during classroom instruction. In order to gather information on this assumption, a video coding manual that allows for analyzing communication processes during chemistry lessons has been developed. Referring to study 1, study 2 is presented to get more information about opportunities students have to develop reasoning and decision-making competences. As a precondition, students' content knowledge is analyzed and especially teachers' dealing with students' conceptions as these situations form the most frequent opportunities for students to learn reasoning skills.

Background

With regard to commonly accepted literature verbal communication during classroom instruction serves as important precondition for students to acquire scientific reasoning skills (Chin, 2006; Edwards & Mercer, 1987; Mehan, 1979; Mortimer & Scott, 2003). In order to explore teachers' and students' classroom discussion and interaction, video studies are a very appropriate tool for a number of reasons (Janik & Seidel, 2009; Pauli, Reusser, & Grob, 2007; Roth et al., 2006; Seidel, Prenzel, & Kobarg, 2005; Stigler, Gallimore, & Hiebert, 2000): Contrary to class registers, which often only render information on the topic and teacher's perspective, videotaped lessons can provide information directly and in an objective manner. Many different aspects can be analyzed in consecutive steps. Thus, a video tape could be analyzed using different categories and new categories could be added, if required. Even information that was possibly overlooked at first glance can be gathered by replaying the tape. Quality and reliability of analysis depend on an adequate coding manual, which includes different criteria followed by descriptive examples on how to classify students' and teachers' statements with regard to the categories.

The TIMSS 1999 Video Study was pivotal in classroom research using video recordings (Roth et al., 2006). In terms of teacher-student communication, the study revealed that, for instance, "whole-class talk [sic] was more likely to take the form of a public presentation (usually by the teacher) than a back and forth public discussion among students and teachers" (Roth et al., 2006, p. 111).

Additionally, differences between five participating countries could be found. In the Czech Republic, a larger percentage of classroom discussion is reported than in Australia, Japan, the Netherlands and the United States (Roth et al., 2006). Varying results in terms of classroom discussion could be found in physics in Germany as well: While no systematic differences between different kinds of teacher and student statements (e.g. percentages of teacher explanations) are reported between the topics of electric circuits and force concept, instructional activities in terms of student experiments seem to vary depending on the topics. While force concept lessons seem to have a teacher-centered instructional organization, the topic of electric circuits is taught in a more student-centered way including more student experiments (Seidel & Prenzel, 2006). A more qualitative approach was chosen by Chin (2006), who explored 14 7th grade science classes of two teachers in Singapore. Following the IRE/IRF concept (Mehan, 1979; Sinclair & Coulthard, 1975), the discourse was analyzed in terms of a triadic dialogue (Lemke, 1990) – teacher initiation, student response and teacher evaluation or follow-up. Chin developed a framework for a “questioning-based discourse” (Chin, 2006, p. 1322) and revealed different ways in which teachers use questions and react to students’ answers. E. g. the follow-up component can have various forms, so a teacher question answered by a student can be followed by a teacher comment or a statement. Additionally, either another question or further statements are posed depending on the purpose of the discourse (e. g. elicit, reply, probe, extend) (Chin, 2006). Typical communication patterns in chemistry lessons had previously been identified via audio lesson tapes (Pitton, 1997). One important result is that teachers mainly ask one-way questions, which are used to guide students to the teachers’ aims. The questions often lead the students’ answers. Qualified and reasonable student statements are not fostered in a whole-class discussion. Along with these results, students’ preconceptions and ideas are not acknowledged and discussed (Sumfleth & Pitton, 1999).

The study described focuses on students’ conceptions and how teachers deal with them in real classroom situations. Best opportunities for interpretation of reciprocal understanding provide situations in which students’ daily-life conceptions (called pre-conceptions) come up in classroom discussion. Because instruction is related to verbal interaction, communication patterns in the classroom are closely analyzed. The basic assumption is that chemistry teachers who offer their students the chance to provide rationale for their statements have the opportunity to learn about their students’ preconceptions. In order to investigate whether students obtain the opportunity to acquire argumentation and communication competences, classroom discussion is analyzed using videotapes of chemistry lessons and a theory-based category system (Pollender & Tepner, 2011b). Thereby, statements are analyzed in terms of correctness and whether students are encouraged to give reasons for their (correct or incorrect) statements (Figure 3).

Methods

The above-mentioned theory-oriented category system was developed based on the standards of content analysis. Referring to Bos and Tarnai (1999), the process of development can be divided into five steps: Theory, development of category systems, training, coding and data analyses, interpreting of results. The development of the category system comprises the operationalization of categories, the formulation of coding instructions with several examples and the selection of evaluation units (time or event sampling). In the third step (training), categories were validated by collaboration of five experts in the field of chemistry education and video analysis. Therefore, 20 videos of grammar school (Gymnasium), grade 7/9/10, and 15 videos of lower secondary school (Hauptschule), grade 9/10, have been analyzed. As a consequence, categories and the coding manual were revised. The manual was developed for analyzing both the surface and deep structures of instruction. While aspects of surface structures are directly observable such as the amount of teachers' and students' speaking time and the length of sentences, aspects on the deep structure level have to be interpreted (e.g., students' cognitive activation). Particularly for analyzing the deep structure, a good coding manual is very important for sufficient rater accordance.

The main focus of the rating was set on students' reasoning and teachers' consciousness of this aspect, so different coding sequences were carried out. Students' and teachers' statements were referred to correctness of students' statements, as it is a criterion for students' understanding and content knowledge. At first, students' statements were classified as correct or incorrect from the chemistry perspective. Secondly, students' reasons for their answers were differentiated between "correctly argued", "incorrectly argued" and "not argued". Thirdly, data was interpreted with regard to the relation between the answer and its foundation. If a correct answer was incorrectly reasoned, the student is thought to have an incorrect understanding, indicating a possible misconception. The same is true if a wrong statement was incorrectly argued. However, if an incorrect statement was justified by a substantiation using the correct chemistry relations or concept, it was interpreted as a simple mistake (e.g., a verbal error). In this case the substantiation would not fit the first statement which, at best, would be revised. If a statement was not argued anyway, no interpretation on a deep structure level was possible (Figure 3). Correct and incorrect statements from the scientific point of view and the corresponding justifications will differ in their "quality" understood as technical terminology, deepness of reasoning or direct way of explanation which will lead to a variety of sub-patterns. Therefore this variety is avoided by deciding between right or wrong in order to end with some main patterns and not to come up with a lot of single cases.

The sample which was taken for the coding manual's development, consisted of 15 videos of chemistry teachers who work at upper secondary schools

(Gymnasium). Seven lessons were videotaped in grade 7 on the topic “state of matter”. Eight videos referred to the topic “acids and bases”, which was taught in grade 10.

Results

As a main result of this study, theory-based communication process diagrams are developed in order to analyze students’ and teachers’ communication structures. These diagrams are useful for presenting communication data in a lucid, graphical way. The x-axis of this two-dimensional coordinate system represents time, while the y-axis shows how communication is progressing. Most of communication processes could be referred to the IRE/IRF concept (Mehan, 1979; Sinclair & Coulthard, 1975). In the first example of the prototypical communication process diagram (see Figure 3, Understanding), a teacher’s answer/impulse/question (“teacher initiation”) is followed by a student’s correct answer (“student response”). The teacher asks the student to justify his or her statement (“teacher follow-up”), and the student responds accordingly. On a deep structure level, it can be concluded that the student has understood the concept.

If the student’s answer is wrong (red ellipse, examples two and three), an incorrect conceptual understanding (example 2) or a simple mistake (example 3) can be assumed. The distinction depends on the student’s explanation. If this is wrong, it might indicate a misconception, while correct reasoning suggests a simple mistake. The presented approach to analyzing communication processes in chemistry education seems to be adequate: Both surface and deep structures are rated clearly.

Figure 3. Communication process diagram, prototype (Pollender & Tepner, 2011a).

Analyses of communication patterns revealed that teachers generally dominated lessons by controlling verbal communication even among the students. One example is shown in Figure 4.

Figure 4. Communication process diagram, example (Pollender & Tepner, 2011a).

To understand the coding of the 10th grade example of a lesson on acids and bases, the transcript of the teacher-student dialogue shown in Figure 4 is presented in translated form (Figure 5). Following legend in Figure 4, “T” means teacher, “S6” means student no. 6, and S3 means student no. 3.

T: Which component do we have to combine with the acids to recognize a change of colours [of the indicator]?

S6: With the base.

T: Which base do you use?

S6: Mmh, what do you call it?

T: It is the most famous one. The most famous base, we already used it to neutralize other acids.

S6: Potash lye.

T: It's potash lye, isn't it?

S3: Sodium hydroxide.

T: Sodium hydroxide. Also called soda lye. In which beaker can we put it to initiate the neutralization process?

Figure 5. Transcript communication processes, example (Pollender & Tepner, 2011a).

According to Pitton (1997), this example seems to be typical of communication in chemistry lessons in Germany. The first analysis of communication patterns in the 10th grade classes gives hints that the course of the lesson is decisively determined by teachers' guidance as the teacher attempts every second communication step. There are no direct student responses to student statements. Both in this example and in further lessons analyzed during the video coding manual development the teacher predominantly asks short questions and the communication process takes place between the teacher and only a few students in a direct way. It seems that teachers ask guiding questions by putting the answer in their students' heads, because students frequently answer in one word and short sentences. According to Chin (2006), Lemke (1990) and Mehan (1979), the discourse structure in these first analyses could be described as a triadic dialogue. After teacher's initiation (e.g. a question, impulse), a student response follows. However, confirmation and corrections (follow-ups) are not

given immediately after students' responses, but after asking other guiding questions. Student-student interaction and classroom discourse barely occur in these examples. In analyzed 10th grade lessons, the teacher did not foster students' reasoning, e.g. by asking for student's explanations or by providing time for classroom discourse. The teacher mainly seems to ask for signal words and declarative knowledge instead of argued contexts. Therefore, students' lack of reasoning skills is not astonishing¹. All in all, the suggested communication process diagrams seem to be an appropriate tool for investigating teachers' and students' communication patterns in future studies. Their biggest advantage, in comparison to other methods of analyzing classroom discourse, is that they retain detailed information of communicative processes and, thus, enable content-specific analyses.

Study 3: Assessing teachers' knowledge

The previous results substantiate the hypothesis that the lack of competences in reasoning and scientific communication might be caused by the predominant student-teacher discussion. Although this relation is only hypothesized and not yet proved, there is good reason to look for parameters which may cause this kind of student-teacher dialogue. One important factor may be deficiencies in the teachers' pedagogical content knowledge, especially with regard to students' preconceptions. Not knowing students' preconceptions and their importance for students' learning processes may cause teacher-dominated communication processes and prevent teachers in chemistry classes from encouraging student explanations and supporting decision-making in complex situations (Van Dijk & Kattmann, 2010). Especially the aspect of pedagogical content knowledge could be related strongly to a deficient content knowledge as recognizing incorrect scientific statements requires a profound scientific knowledge. So main purpose of this study is assessing teachers' CK and PCK. Therefore, different PCK facets (knowledge about students' preconceptions, knowledge about models and experiments) are compared to each other. Additionally, knowledge of teachers working at non-intensified schools is compared with knowledge of teachers who work at intensified schools, because students on non-intensified level may bring in more inadequate preconceptions than students on intensified level.

Background

Professional knowledge is considered fundamental for successful teaching (Abell, 2007; Peterson, Carpenter, & Fennema, 1989). However, in literature, there is no consensus regarding the construct and the interaction of single dimensions of professional knowledge. Referring to Shulman (1987), three

¹ Additionally, the teacher did not elaborate on the difference between (solid) sodium hydroxide and (aqueous) soda lye.

aspects can be seen as common characteristics of all known definitions of professional knowledge (Fischer, Borowski, & Tepner, 2012): CK, PCK, and PK (pedagogical knowledge). CK is conceptualized as knowledge of subject-specific contents and methods and is assumed to be a prerequisite for developing PCK that is necessary to transform subject knowledge into school practice. The non-subject specific dimension of PK is knowledge about broad principles of classroom management, organization and learning strategies (Baumert & Kunter, 2006). PCK could be understood as an overlap between general pedagogical knowledge (PK) and subject-specific content knowledge (CK). Because CK is related to different topics, both CK (as a precondition for PCK) and PCK vary across different themes taught in school especially with regard to students' preconceptions. So research on CK and PCK should focus on the same topics. By choosing topics that are taught in almost each grade like e.g. acids and bases, generalization of studies' findings could be fostered.

The construct of PCK could further be divided into several "Conceptions of Purpose for Teaching Subject Matter" (Grossmann, 1990, p. 5) or facets/components (Park & Oliver, 2008). Following Shulman (1986), most scholars agree with two core elements of PCK: "(a) knowledge of instructional strategies incorporating representations of subject matter and understanding of specific learning difficulties and (b) student conceptions with respect to that subject matter" (Park, Jang, Chen, & Jung, 2011, p. 248). In chemistry knowledge of instructional strategies could be interpreted as knowledge of dealing with models and knowledge of dealing with experiments. Both aspects refer to teachers' knowledge whether a model (plastic or theoretical) or experiment fits a certain topic in a special situation and if it is useful to facilitate students' understanding. In this study, students' conceptions are conceptualized both as notions correct from a scientific point of view and misconceptions that are scientifically wrong. Students come to class having implicit theories that are often incorrect or incomplete (Kuhn, 2002). As a first step to overcome this situation, students have to "acknowledge that one's existing knowledge is incomplete, possibly incorrect" (Kuhn, 2002, p. 373). Both teachers' and students' questions can help scrutinize students' knowledge. In order to make classroom discussions more fruitful to foster conceptual change, teachers should be aware of correct and incorrect preconceptions of their students. Teachers' knowledge about students' preconceptions could be compared with knowledge about models and experiments to gather information on relative difficulty of these PCK facets and on the influence of CK on PCK facets. It can be assumed that knowledge about students' preconceptions requires more CK than knowledge about models and experiments.

There is a lack of knowledge concerning the influence of teachers' CK and PCK on their behavior in chemistry lessons (Abell, 2007), because no adequate test instruments for assessing both teachers' professional knowledge and their classroom actions including organization of classroom discourse exist to date.

Supplementary to the theory-oriented category system for analyzing classroom activities described above, an adequate test instrument for assessing teachers' knowledge is required, as from a researcher's point of view, gathering quantitative data samples is necessary in order to calculate correlations and to draw generalizable conclusions. While instruments for measuring content knowledge could be based on long-term experiences in measuring students' knowledge, the development of tests for measuring pedagogical content knowledge is especially difficult. So, developing items for a CK test is comparatively easy, while teachers' PCK is more difficult to explore in particular because there may be no decisively right or completely wrong answers. Therefore, a test instrument was used that covers certain topics and that can be used for triangulation with video data.

The development of a large-scale test instrument for assessing chemistry teachers' CK and PCK was the main aim of a project which deals with science teachers' knowledge (Witner & Tepner, 2010). With regard to communication and argumentation competences, however, it has to be explored if teachers' PCK on students' preconceptions plays a particular role within the researched PCK facets. Correlations of PCK with CK and between different aspects of PCK were therefore calculated and compared.

Methods

For gathering information about chemistry teachers' CK and PCK on students' preconceptions, paper-pencil tests were used in a closed item format (Witner & Tepner, 2010). Because PCK is bound to certain teaching situations and CK is assumed to be a prerequisite for the development of PCK, the tests focused on the same content. As topics, "structure of atoms and the periodic table of the elements", "chemical reactions using the example of acids and bases", and "chemical bonding" were chosen. Both tests included items covering different knowledge areas: Declarative knowledge ("knowing that"), conditional knowledge ("knowing when and why"), and procedural knowledge ("knowing how") (Paris, Lipson, & Wixson, 1983). PCK items assess teachers' knowledge about models, experiments, and students' preconceptions which form the core of this article. The facets 'models' and 'experiments' were added as it could be assumed that teachers performing poorly in these two facets behave in a less flexible way to students' answers than teachers who perform strongly. Both knowledge of models and knowledge of experiments can be used to help students overcome their misconceptions. Additionally, these two facets also appear as comparison to students' preconception.

Analysis of an unknown substance has revealed following results:

- 1. State of aggregation: solid**
- 2. Melting point: 782 °C**
- 3. Solubility in water: 740 g/L**
- 4. Solution conducts electric current.**

Which statement can be definitely deduced regarding the unknown substance?

Check your answer! Only one answer is correct.

☐ This substance is deformable.

☐ Inside this molecule there is a polar bond towards a hydrogen atom, and there are free electron pairs.

☒ Molten mass of this substance conducts electric current.

☐ The difference regarding the electro negativity of the elements is at most 1.0.

Figure 6. Example for a CK test item.

While the CK test (Figure 6) consisted of 29 items in a multiple choice/single select format (four alternatives each), the PCK test had 20 items. Nine items referred to students' preconceptions (Figure 7).

In a middle school lesson of your trainee teacher, a student explains the increase in volume during the process of heating using following words:

"Substances expand when being heated because of the expansion of their particles."

Please assess following possible continuations with regard to students' misconceptions. Please assign marks from 1 ("very good") to 6 ("unsatisfactory") to each single alternative.

	1	2	3	4	5	6
a) Students simulate states of aggregation (each student represents one particle). So they will realize that they will need more space, when they move.	☐	☐	☐	☐	☐	☐
b) Approval of the statement's first claim by the teacher, correction of the second claim.	☐	☐	☐	☐	☐	☐
c) Checking the hypothesis by melting a specific volume of ice.	☐	☐	☐	☐	☐	☐
d) Motion of particles will be chosen as the subsequent focal topic.	☐	☐	☐	☐	☐	☐

Figure 7. Example for a PCK test item.

Each PCK item comprised a description of a certain classroom situation and four alternatives how to react adequately. Chemistry teachers were asked to assess the adequacy of each alternative by a six point rating scale using grades from 1

= “very good” to 6 = “unsatisfactory”. Results were analyzed using quasi-ranked pairs (Thillmann, 2008). This means, two answer alternatives were compared to each other at a time. It is not the exact numerical value that is relevant for scoring, but the relation between two alternatives (e.g., alternative “a” is better than “c”). To identify a point of reference for the teacher scoring, an expert rating was conducted beforehand. Thirteen experts in the field of research on chemistry education and heads of practical teacher training seminars in chemistry were asked to assess each alternative. As a high reliability ($\alpha = .91$) between experts could be found, all of them were included in the analysis. Teachers were only awarded one point if two answer alternatives were related in the same way the experts rated them. For exploring teachers’ PCK concerning students’ preconceptions, 26 relations covering nine items were taken into account. The whole PCK test referred to 65 relations distributed over 19 items. Processing time was about 45 minutes in total. This item format was chosen to deal with the problem of finding PCK answer alternatives that were decisively right or totally wrong (Dollny, 2011). All in all, 166 teachers took part voluntarily in this study.

Results

Data referring to the pilot study show that both the reliabilities of the CK test ($\alpha = .84$) and of the whole PCK test were good ($\alpha = .83$). The reliability of the subscale on students’ preconceptions was sufficient ($\alpha = .79$). The special research question focused on correlations between PCK on students’ preconceptions and PCK on models and experiments, as well as CK. Data analysis of 166 teachers (pilot and main study) revealed significant correlations (Table 2). While all correlations were on a medium level (Field, 2009), the correlation between PCK on students’ preconceptions and CK was shown to be stronger than the correlation of PCK with regard to the sub-areas students’ preconceptions, models, and experiments. Possibly, CK could be a necessary precondition especially for knowledge of and dealing with students’ preconceptions. As first results show, a profound knowledge base in subject matter seems to be important to understand learning difficulties. Yet, further studies have to be conducted.

Table 2. *Correlations between PCK on students’ preconceptions and PCK on models and experiments, and CK.*

Correlated Topic/Dimension	<i>r</i> (Spearman)	<i>p</i>
PCK on models	.286	< .001
PCK on experiments	.284	< .001
CK	.415	< .001

A descriptive comparison of the mean scores of the teachers shows that answering items concerning the PCK aspect of students’ preconceptions seems

to be most difficult (Figure 8). It is significantly more difficult than answering items concerning PCK on experiments ($z = -4.500$; $p < .001$; $r = -0.441$) and PCK on models ($z = -5.984$; $p < .001$; $r = -0.587$).

Figure 8. Results – PCK on students' preconceptions in comparison to the other facets.

Additionally, knowledge about student preconceptions seems to depend on the school type at which the teachers work. Teachers who work at upper secondary school (Gymnasium) score better than teachers who work at schools of the basic general education level (Hauptschule, Realschule), $t(163) = 4.33$; $p < .001$; $d = 0.81$ (Figure 9).

Figure 9. Results – PCK on students' preconceptions in comparison of two different school types.

Although it could be assumed that students at non-intensified schools bring in more inadequate preconceptions, teachers who work at these schools have less knowledge of their students' preconceptions. It is possible that instruction is more teacher-centered, so there are fewer opportunities for students to bring in their conceptions.

Furthermore, PCK on students' preconceptions in chemistry seems to be special for chemistry teachers, at least with regard to topics investigated.

Compared to a group of biology and physics teachers, it was found that chemists ($t(52) = 2.63$; $p = .011$; $d = 0.77$) and chemistry teachers ($t(200) = 5.43$; $p < .001$; $d = 1.00$) scored better (Figure 10). The measured constructs seem to be special and not part of a general subject unspecific knowledge. Furthermore, chemistry teachers scored significantly better than biology and physics teachers. Chemists and chemistry teachers were not found to differ significantly. This might be caused by the important influence of content knowledge in this sub-area of PCK.

Figure 10. Results – PCK on students' preconceptions in comparison of different groups of science teachers and chemistry literate persons.

All in all, the used test instruments seem to measure chemistry teachers' CK and PCK related to the subscale of students' preconceptions. Low correlations with 'models' and 'experiments' indicate distinct PCK facets. In addition, CK and PCK on students' preconceptions seem to be different dimensions of professional knowledge. They are correlated with each other on a medium level.

Discussion

The results of the first study showed that students' decision-making skills depend on the content in question. These competences are more developed in everyday-life situations than in scientific situations, especially with regard to chemistry content. While students showed that they were able to use adequate decision-making strategies in everyday-life situations, they lack judgment skills in SSI. The students scored significantly higher in test items on everyday-life situations compared to scientific situations, although the items were constructed as similar as possible. Additionally, the analysis of the data showed that the three different contexts (biology, chemistry, everyday-life) are distinguishable

latent constructs, supporting the assumption that different skills are needed to make decisions in these contexts. One reason for the lack of competence in scientific reasoning might be that decision-making plays a less important role in science lessons. This would mean that the students possess the basic skills to make justified decisions but are not experienced in applying these skills to scientific problems. On the other hand, they are used to apply these skills in everyday-life situations very often (planning holidays, buying a mobile phone, choosing products, etc.) which makes it easy for them to apply these skills to non-scientific situations. These findings result in the hypothesis that in science lessons, students are used to learning and understanding the content and to conducting experiments, but not to arguing on (socio-)scientific issues.

In order to investigate whether the reason for the deficits can be found in the design of science lessons, an instrument for analyzing the course of a lesson is required. The first indications in support of this hypothesis can be derived from the second study. In order to analyze students' and teachers' communication, a video coding system was developed. It is able to depict communication patterns in science lessons with a focus on teachers' reactions to students' statements. The category system is able to describe in which way teachers try to trigger students' argumentation or not. Comparable to the TIMSS video study, the analysis of the communication patterns shows a poor performance of German teachers in encouraging argumentation processes. The teachers mainly ask questions that can be answered quickly and do not particularly activate student-student discussions (e.g., on students' preconceptions or on SSI).

For this reason, German students have only few opportunities to practice scientific reasoning at school, and there may be different reasons that explain this result. One could be that the curricula focus on content knowledge to a high degree, although this is not true for the current German curricula. Another explanation is that the teachers themselves are not able to advise students in scientific reasoning because this topic is quite new in Germany and was not necessarily part of teacher education in the last decades. Since a correlation of teachers' PCK and teachers' organization of classroom discourse can be assumed, a test instrument was used in the third study to measure different facets of CK and PCK. With regard to PCK, the main aspects measured are knowledge of students' preconceptions, models and experiments. The underlying idea was that teachers are especially able to foster students' argumentations if they are able to identify inconsistent conceptions. If these conceptions are identified by the teacher, they can be used as a starting point for a scientific discussion. The results of the PCK test show a low score for knowledge of students' preconceptions compared to the other two facets, "models" and "experiments". Additionally, the correlation between PCK on preconceptions and CK is the highest found in the analysis.

The next step will be to administer these tests to teachers who are taking part in video studies and to try to find correlations between special features described

by the categories above and teachers' PCK on students' preconceptions. Therefore, a system for scoring the video data is being developed at the moment. The results of students' performance, teachers' performance and organization of classroom discourse can then be triangulated in order to gain a deeper insight into preconditions for fostering students' abilities in decision-making and scientific reasoning. Intervention studies in the field of in-service teacher training courses may then help to determine the effective direction. In the long run, the aim is to optimize both students' and teachers' competences – both in teaching and in learning effective decision-making and reasoning techniques.

References

- AAAS Project 2061. (2001). *Atlas of science literacy*. New York / Oxford: Oxford University Press.
- AAAS Project 2061. (2007). *Communicating and learning about global climate change. An abbreviated guide for teaching climate change, from Project 2061 at AAAS*. Washington: AAAS Publication Services.
- Abell, S. K. (2007). Research on science teachers' knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Aufschnaiter, C. von, Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131. doi:10.1002/tea.20213
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Bos, W. & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, 31, 659–671.
- Bybee, R. W. (2002). Scientific Literacy - Mythos oder realität? In W. Gräber, P. Nentwig, T. Koballa, & R. Evans (Eds.), *Scientific literacy* (pp. 21–43). Opladen: Leske + Budrich.
- Bybee, R. W., Fensham, P. J., & Laurie, R. (2009). Scientific literacy and contexts in PISA 2006 Science. *Journal of Research in Science Teaching*, 46(8), 862–864.
- Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, 28(11), 1315–1346.
- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37(6), 582–601. doi:10.1002/1098-2736(200008)37:6
- Dollny, S. (2011). *Entwicklung und evaluation eines testinstruments zur erfassung des fachspezifischen professionswissens von chemielehrkräften* (Dissertation). Universität Duisburg-Essen.
- Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London, UK: Heinemann.
- Eggert, S. & Bögeholz, S. (2010). Students' use of decision making strategies with regard to socioscientific issues - An application of the rasch partial credit model. *Science Education*, (94), 230–258.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.
- Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 435–448). New York: Springer.

- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J. et al. (Eds.). (2009). *PISA-2006-Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Münster, New York, NY, München, Berlin: Waxmann. Retrieved from <http://d-nb.info/994990073/04>
- Göbel, J. & Walpuski, M. (2010). Modelling students' evaluation and judgement competence in chemistry education considering external influential aspects. In I. Maciejowska & P. Ciesla (Eds.), *10th european conference on research in chemistry education* (pp. 104–105). Krakow: Pedagogical University of Krakow, Institute of Biology, Department of Chemistry and Chemistry Education.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education. Professional development and practice series*. New York: Teachers College Press.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeits-Test für 5.-12./13. Klassen, Revision - Materialien-Manual-Koffer*. Göttingen: Beltz Test.
- Hostenbach, J. (2011). *Entwicklung und prüfung eines modells zur beschreibung der bewertungskompetenz im chemieunterricht*. Essen: Universität Duisburg-Essen.
- Janik, T. & Seidel, T. (2009). *The power of video studies in investigating teaching and learning in the classroom*. Münster: Waxmann.
- Kauertz, A. & Fischer, H. E. (2006). Assessing students' level of knowledge and analysing the reasons for learning difficulties in physics by Rasch analysis. In X. Liu & W. J. Boone (Eds.), *Applications of Rasch measurement in science education* (pp. 212–246). Maple Grove, USA: JAM Press.
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene kompetenzmodellierung in den naturwissenschaften der sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135–153.
- KMK. (2005). *Bildungsstandards im fach chemie für den mittleren schulabschluss*. Neuwied: Luchterhand.
- Kortland, K. (1996). An STS case study about students' decision making on the waste issue. *Science Education*, 80, 673–689.
- Kuha, J. (2004). AIC and BIC: comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2), 188–229. doi:10.1177/0049124103262065
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 371–393). Oxford: Blackwell Publishing.
- Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84(1), 71–94. doi:10.1002/(SICI)1098-237X(200001)84:1<71::AID-SCE6>3.0.CO;2-C
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, N.J: Ablex Pub. Corp.
- Mehan, H. (1979). *Learning lessons*. Cambridge, MA: Harvard University Press.
- Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead, UK: Open University Press.
- National Research Council. (1996). *National science education standards*. Washington, D.C.: National Academic Press.
- Neumann, K., Fischer, H., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*, 8, 545–563. Retrieved from <http://dx.doi.org/10.1007/s10763-010-9206-7>
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293–316.
- Park, S., & Oliver, S. J. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.

- Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, 41(2), 245–260. doi:10.1007/s11165-009-9163-8
- Posada, David; Buckley, Thomas (2004): Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. In: *Systematic Biology* 53 (5), S. 793–808.
- Paulhus, D. L. (1998). *The balanced inventory of desirable responding*. Toronto/Buffalo: Multi-Health Systems.
- Pauli, C., Reusser, K., & Grob, U. (2007). Teaching for understanding and/or self-regulated learning? A video-based analysis of reform-oriented mathematics instruction in Switzerland. *International Journal of Educational Research*, 46(5), 294–305.
- Peterson, P. L., Carpenter, T. P., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlating and case analysis. *Journal of Educational Psychology*, 81(4), 558–569.
- Pitton, A. (1997). *Sprachliche kommunikation im chemieunterricht. eine untersuchung ihrer bedeutung für lern- und problemlöseprozesse. naturwissenschaften und technik - didaktik im gespräch*. Münster: Lit.
- Pollender, T. & Tepner, O. (2011a). *Using communication patterns for analyzing classroom activities in chemistry: Poster*. Lyon: ESERA 2011 Conference.
- Pollender, T. & Tepner, O. (2011b). PCK von chemielehrkräften – Entwicklung eines videokategoriensystems. In D. Höttecke (Ed.), *Naturwissenschaftliche bildung als beitrage zur gestaltung partizipativer demokratie. Gesellschaft für didaktik der chemie und physik. Jahrestagung in Potsdam 2010* (pp. 537–539). Berlin: Lit.
- Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T. et al. (2006). *Teaching science in five countries: Results from the TIMSS 1999 video study statistical analysis report*. (NCES 2006-011). U.S. Department of Education, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office. Retrieved from <http://nces.ed.gov>.
- Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, 41(5), 513–536.
- Sadler, T. D. & Zeidler, D. L. (2005a). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching*, 42(1), 112–138.
- Sadler, T. D. & Zeidler, D. L. (2005b). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, 85, 71–93.
- Seidel, T. & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, 16, 228–240.
- Seidel, T., Prenzel, M., & Kobarg, M. (2005). *How to run a video study*. Münster: Waxmann.
- Sharan, S. (1980). Cooperative learning in small groups. *Review of Educational Research*, 50, 241–271.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching of the new reform. *Harvard Educational Review*, 57, 1–22.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87–100.

- Sumfleth, E. & Pitton, A. (1999). Problems of communication in chemistry classes. In M. Bandiera, S. Caravita, E. Torracca, & M. Vicentini (Eds.), *Research in science education in europe* (pp. 73–80). Dordrecht: Kluwer.
- Thillmann, H. (2008). *Selbstreguliertes lernen durch experimentieren: Von der erfassung zur förderung*. Universität Duisburg-Essen. Retrieved from <http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-18970.xml>
- Uskola, A., Maguregi, G., & Jiménez-Aleixandre, M.-P. (2010). The use of criteria in argumentation and the construction of environmental concepts: a university case study. *International Journal of Science Education*, 32(17), 2311–2333.
- Van Dijk, E. M. & Kattmann, U. (2010). Evaluation im unterricht: eine studie über fachdidaktisches wissen von lehrerinnen und lehrern. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 7–21.
- Walpuski, M., Ropohl, M., & Sumfleth, E. (2011). Students' knowledge about chemical reactions - development and analysis of standard-based test items. *Chemistry Education Research and Practice*, (12), 174–183.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Göttingen: Hogrefe & Huber.
- Witner, S. & Tepner, O. (2010). Professional knowledge of chemistry teachers – test development and evaluation. In M. F. Taşar & C. Çakmakci (Eds.), *Contemporary science education research: preservice and inservice teacher education* (pp. 223–228). Ankara, Turkey: Pegem Akademi.