Research on Educational Standards in German Science Education – Towards a model of student competences

Christoph Kulgemeyer & Horst Schecker
University of Bremen, GERMANY

This paper gives an overview of research on modelling science competence in German science education. Since the first national German educational standards for physics, chemistry and biology education were released in 2004 research projects dealing with competences have become prominent strands. Most of this research is about the structure of science competence as laid out in the standards. We first discuss the notion of competence in general and of science competence in particular. We then present a selection of results with respect to competence modelling. Finally, we critically review the impact of this research on teaching and show perspectives for future studies and classroom practise.

Keywords: competence, competence modelling, educational standards

INTRODUCTION

The German situation after the first PISA-study

Since the results of the first PISA-Study (Baumert et al., 2001) were published which revealed an unexpectedly low performance of the German students, research on competences and educational standards has become one of the main fields of work of German science education. The results of PISA were dissatisfying for educators, researchers, and the general public. German students only achieved levels below the international average in all three test domains (reading comprehension, mathematics, and science). These mediocre results started a broad discussion about the quality of the German educational system, also in mainstream media. Some researchers refer to this situation as the 'PISA-shock' (e.g., Schecker & Parchmann, 2006).

One of the most important reactions to PISA was the development of educational standards, including the three sciences biology, chemistry, and physics (c.f. Neumann, Kauertz & Fischer, 2010). A major intention was to change the German teaching tradition of detailed curricula and prescriptions of content to be taught (input orientation). So far, each of the 16 federal states in Germany had decided independently about their own curricula. The content could differ considerably between the states. The new German National Educational Standards (NES) (KMK, 2005) follow a different approach. Instead of describing general aims and science content, the standards are devised as achievement standards, i.e. as the knowledge and abilities an average student is expected to have developed after 10 years of school (outcome orientation). "An average student" in this respect can be described as a student who attended the regular number of science classes and has thus gone through the regular syllabus with medium results in performance tests. The NES were a joint decision of all the federal states. The curricula of the states from then on had to accord with

Correspondence to: Christoph Kulgemeyer,
Institute of Science Education, Department of
Physics Education, University of Bremen, Otto-
Hahn-Allee 1, 28359 Bremen, Germany
E-mail: kulgemeyer@physik.uni-bremen.de
doi: 10.12973/eurasia.2014.1081a

State of the literature

- As a consequence of Germany's mediocre results in international student performance studies like PISA, the German educational system has turned from input-orientation towards output-orientation. In 2004 national standards for the sciences were published
- Since the standards need evaluation, research on assessing competences has been a major field of German science education research in the last decade
- Most of this research focuses on modelling students' competences in a research perspective. These models have been validated with a large empirical effort. There are, in contrast, only few projects developing and evaluating competence models and according curricular materials for teaching purposes

Contribution of this paper to the literature

- We show the relationships between general considerations about the concept of competence, science-specific competences, and the German national education standards. Against this background we explain why the concept of competence is so widely used in the current educational discourse on aims and standards in Germany
- We contrast the research that deals with the evaluation of the standards with alternative frameworks and present a brief overview of the key results of this research
- Our paper serves as a window to an understanding of the process of competence modelling in Germany. We also name implications for further research, such as an empirically guided development of materials to teach competences in science

the NES. The purpose of the curricula is to specify and operationalize the goals of science education. Operationalization and specification are supposed to help to ensure compliance of the standards. For the evaluation of the standards the federal states founded a scientific institute for quality development in education (Institut zur Qualitätsentwicklung im Bildungswesen, IBQ, Berlin).

In this paper, the concept of competence will be described. Afterwards, we will provide insights into the development of competence models especially with respect to the German NES. This paper holds a critical view on the recent process of standards evaluation: We will finish with a look on alternative frameworks and the impact of this part of German science education research on actual science teaching.

The concept of competence in science education

Over the years, the term 'competence' has been discussed not only in science education literature but also in pedagogy and psychology. As a result, there are several different notions of competence. The idea of competence has been established in pedagogy and psychology in order to describe a disposition that enables a person to perform successfully in content-related, complex and demanding problem situations. Science competence for example can be understood as a person's latent trait to successfully solve science tasks, like developing a setup for an experimental investigation or modelling an everyday-life phenomenon with physics laws and principles. The concept of competence once was developed in psychology because more and more researchers doubted that either factual knowledge or intelligence are sufficient concepts to predict successful actions in a particular challenge (McClelland, 1973). Shavelson (2010) identified six facets from the literature that nearly all of the different notions of competence share: "Competence

is a physical or intellectual ability, skill or both;
is a performance capacity to do as well as to know;
is carried out under standardized conditions;
is judged by some level or standard of performance as 'adequate', 'sufficient', […];
can be improved;
draws upon an underlying complex ability; and
needs to be observed in real-life situations" (Shavelson, 2010, p. 44).

In education, the term became more and more important because it complemented the idea of qualification for a certain profession. Competences were often phrased and understood as general traits of a person, not necessarily specific for a particular domain such as science, or at least easily transferable. This notion is referred to as 'key competencies' (e.g. Maag Merki, 2003). Klafki in contrary used the concept prescriptively in order to describe what enables an individual to solve domain-specific problems. He theoretically worked out two necessary aspects: the required skills and abilities and the required motivation to apply these skills and abilities. (Klafki, 2006). Competence in this understanding is domain-specific. There is evidence to support this opinion: Some competences, which were thought to be general, had to be differentiated between domains. For example problem solving competence was empirically shown to depend on the domain in which it was applied. For physics, it shows a strong relationship with content knowledge (correlation r=0.81; Friege & Lind, 2003, 70). But there are also contradictory approaches to problem solving: In PISA 2003 problem solving was treated under a general perspective (Dossey et al., 2004). Referring to data from PISA 2003 the latent correlations

between problem solving of everyday problems and science, mathematics and reading skills range between .80 and .89 (Dossey et al., 2004, p. 55). Such strong correlations to different other constructs imply a more general ability. If it was for example science specific one would expect a high correlation with science skills (convergent validity) but a lower correlation with the other non-science traits (discriminant validity). However, it is widely accepted that the content or context plays an important role for describing competence (c.f. Kauertz et al., 2012) – the PISA consortium chose to use everyday problems which do not necessarily refer to the curriculum (Dossey et al., 2004, p. 27).

For the German Educational Standards and all the research in this context Weinert describes the underlying concept of competence. He states that competences are "clusters of cognitive prerequisites that must be available for an individual to perform well in a particular content area" (Weinert, 2001, p. 47). This notion on competence is clearly domain-specific as it refers to a particular content area. As a result, the research conducted with reference to the NES deals with identifying, science-specific aspects of competence. Just as the educationalist Klafki, the psychologist Weinert highlights the importance of motivation and willingness to actually apply the cognitive prerequisites in a particular situation. His concept of competence explicitly comprises both of these aspects: ability and willingness. However, in actual research most projects focus on the cognitive aspect of competence, which can be assessed much more easily.

For a systematic description of the competences needed to master problems in a particular domain competence models are specified. They provide the basis for test development and measurement of competences. In the context of educational standards these measurement-related issues of structuring competence have become very important.

Competence modeling

There are different ways to model competences. Schecker & Parchmann (2007) presented two dichotomies that can be useful to characterize such models. Competence models

can either model the structure of a competence (structure models) or the development of a competence (developmental models);
can either be based on normative considerations (normative models) or on empirical evidence (empirical models).

To speak with Weinert's definition of competence, structure models cluster groups of 'cognitive prerequisites' for solving problems in a specific domain. The structure can for example consider different contents (e.g. mechanics or thermodynamics for physics

competence) or different cognitive processes (e.g. reproducing information or selecting relevant information) that are required to solve a task (Kauertz et al., 2012). Such areas (clusters) of abilities (and skills) together form a competence structure. Abilities within an area (like reproducing, processing and transferring science information) can have three different relationships (Einhaus, 2007): they can be totally independent from another, they can be abilities that influence one another in a way that improving one ability leads to an improvement in the other ability, and they can form levels of competence. In the latter case, the one ability is a prerequisite for the other.

Developmental models describe how competences grow and which stages have to be passed to reach higher levels. One could say that structure models are synchronic models and developmental models are diachronic models. A good example for the developmental view is the model of learning progression in energy by Neumann et al. (2013). Even though these authors do not use the term competence, their understanding of learning progressions shows a strong relationship to the development of competence: „[…] Learning progressions not only involve knowledge, but also the abilities and skills required to solve real-life problems […]" (Neumann et al., 2013, p. 164). The importance of a developmental perspective has been emphasized for learning in general (e.g. NRC (2007)). Well-developed developmental competence models are still rare; they need a large empirical effort. But e.g. Neumann et al. (2013) show that also competence modeling considers this perspective more and more explicitly. This might once provide important information for teaching competences more effectively.

These basic kinds of models – structure models and developmental models – make certain assumptions, either about the inner structure of the competence or about its development. These assumptions have to be reasoned. In the case of empirical models, support for a certain inner structure of a competence, e.g. physics competence, is given by empirical data on students' performances in solving tasks and problems. In the case of normative models, theoretical and prescriptive considerations are used. Normative models are especially important with respect to the development of educational standards. Normative considerations lead to expected or, rather, desired structures or developments: The expected results of ten years of school cannot be described without normative deliberations. Empirical models in contrast describe a structure or a development based on empirical evidence. This approach is useful to explain the performances of persons in a test, but not necessarily based on a theory about the structure of their cognitive abilities.

A good example for an empirical model is the model by Rost et al. (2005). They analysed a national German

study accompanying PISA 2003 and found that most normative models overestimated the importance of content areas for the structure of science competence. Instead, content areas only played a minor role for explaining the variance in students' test performances. Cognitive abilities like working with mental models or working with numbers turned out to be more important. A good example for a normative model is the initial German NES as described below.

Basically, competence models can have three different purposes (Klieme et al., 2003):

(1) they can represent the cognitive structure or the mental model an individual holds about a certain domain,

(2) they can describe the relationship of different domain-specific educational aims and make them more concrete for research, and

(3) they can provide orientation for actual science teaching.

The latter purpose is the most important use of competence models for education: to make abstract educational aims more concrete for teachers and teaching. A good means to illustrate educational aims is to present tasks used in competence tests. By giving and characterizing tasks, domain-specific competences and their structures can be illustrated. Many competence models, however, are formulated in a way that mainly fulfils research functions (2). Competence models are often much too sophisticated for teachers to implement them in their actual teaching. They have to be adapted to teaching purposes (Maiseyenka, Schecker & Nawrath, 2013).

The German educational standards

Educational standards are established in several countries, among them the USA (AAAS, 1991), the UK (QCA, 1999), and Australia (MCEETYA, 2005). They all have specific structures. The separate German NES for physics, chemistry, and biology differentiate between four areas of competence: use of science content knowledge, application of epistemological and methodological knowledge, science communication, and judgement. 'Judgement' can be seen as close to decision-making and argumentation in the context of socio-scientific issues.

The standards in these four areas of competence are described as abilities an average student is expected to have achieved at the end of lower secondary education (age 15). Besides the dimension "competence area" there are two more elements of structure in the standards (c.f. Figure 1):

Basic concepts: Organisation of the content knowledge by four core ideas. System, matter, interaction, and energy are assumed to be the basic concepts for physics. The use of the basic concepts in teaching is meant to support a more coherent organisation of the students' knowledge. Basic concepts show structural relationships between science phenomena in different contents. Whether the basic concepts in physics are chosen appropriately for this purpose is a matter of discussion (e.g. Schecker & Wiesner, 2013).

Demands: reproduction, application, and transfer of knowledge

The demands are often interpreted as levels of competence. Two thirds of the standards booklet (KMK 2005) is made up of sample tasks. Each of the tasks is characterized by three parameters: competence area, challenge and basic concept. This leads to a three-dimensional structure for describing the expected competences in the NES (c.f. Figure 1). Every problem that can be solved with science competence refers to a specific component of each dimension. Vice versa, the abilities to master this particular problem can be described with three components. An example would be
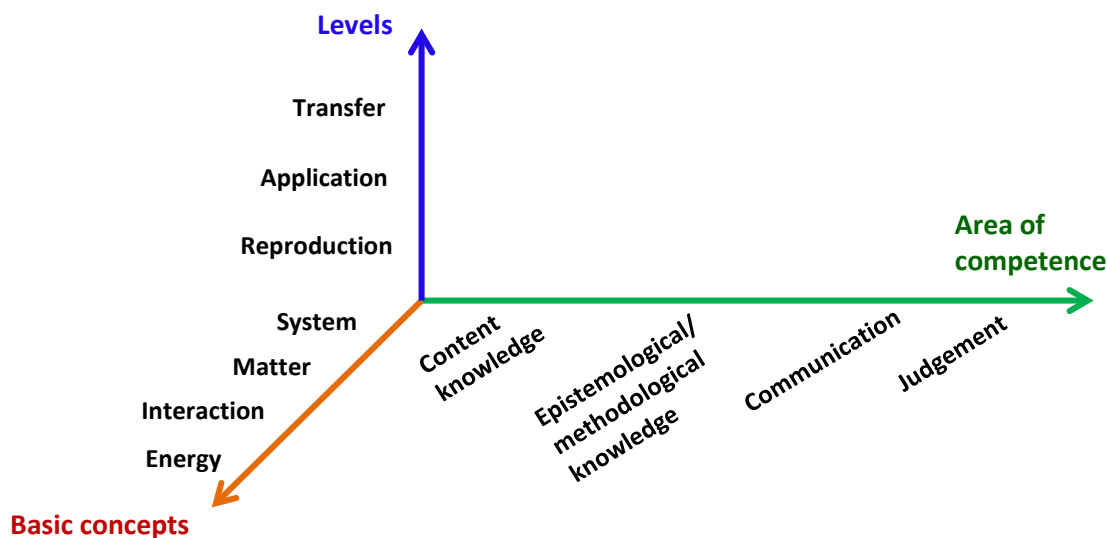

Figure 1. Representation of the underlying competence model used in the German National Educational Standards for Physics (translation by the authors)

a problem that needs the reproduction (dimension 'demands') of content knowledge (dimension 'area of competence') with respect to energy (dimension 'basic concepts'). More explicitly, the calculation of the kinetic energy of a car (mass: 1 t) driving with 20 m/s could be such a task. Another example is the application ('demand') of judgement strategies ('area') to come to a decision in the field of energy saving ('basic concept').

The German NES contain sample tasks to illustrate the three dimensions. The purpose of these tasks is to help teachers find access to the standards.

Evaluation of the German NES

The project *ESNaS (Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I: evaluation of the national educational standards for natural sciences at the lower secondary level)* is a long-term project for test development and evaluation of the NES. The IQB commissioned a group of science education researchers, mainly from the University of Duisburg-Essen, with test development. As the standards leave room for interpretation (Kauertz, Fischer & Siegle, 2013), e.g. referring to the role of the basic concepts (see above), it was difficult to construct appropriate test items. Actually, being not explicit enough for assessment purposes is a problem that concerns every standard or performance expectation (Pellegrino, 2013, p. 320). Major problems arise from the fact that the standards do not contain a national curriculum. Furthermore, the German standards are formulated as so called 'regular standards'. This means they contain the learning outcomes an *average* student is expected to have attained – and it is an empirical problem what an average student actually is. In contrary, e.g. the Swiss standards are basic standards; they formulate the abilities that every student *must* have reached (the minimum).

Specification of German NES model for evaluation purposes

The first step in ESNaS had to be a specification of the competence model (Kauertz et al., 2010) for research purposes. First of all, the dimension 'basic concepts' was applied to specify the *area of competence* 'content knowledge'. The task pool contains items developed for all the basic concepts.

The NES-dimension *demands* was not constructed to describe a hierarchical graduation of a competence but as an orientation for teachers about different types of challenges for students. There is evidence that the demands are not useful for the empirical graduation of item difficulty (Schmidt, 2008). ESNaS replaced the dimension 'demands' by two dimensions 'complexity' and 'cognitive processes' (c.f. Fig. 2). Both had shown to be helpful to operationalize item difficulty in prior studies (e.g. Commons et al., 2007; Kauertz, 2008; Bernholt, 2010).

The dimension *cognitive processes* refers to processing given information for the solution of a task (reproduce, select, organize, integrate). The influence of cognitive processes on the difficulty of a task is a well-known issue in psychology and science education as well (Kremer et al., 2012; Adkinson & Shiffrin, 1968). *Reproduction* means that facts, relations or concepts simply need to be recalled from a given representation form like a text or a diagram (or from prior knowledge). *Selection* means that one has to decide which facts, relations or concepts in a given set of information is relevant for solving a specific task; e.g. from information represented in a diagram. *Organisation* refers to the need to give facts, relations or concepts a structure; e.g. term transformations with given formulas would need the process of organisation. Finally, *integration* means that connections between given pieces of information have to be worked out connected with
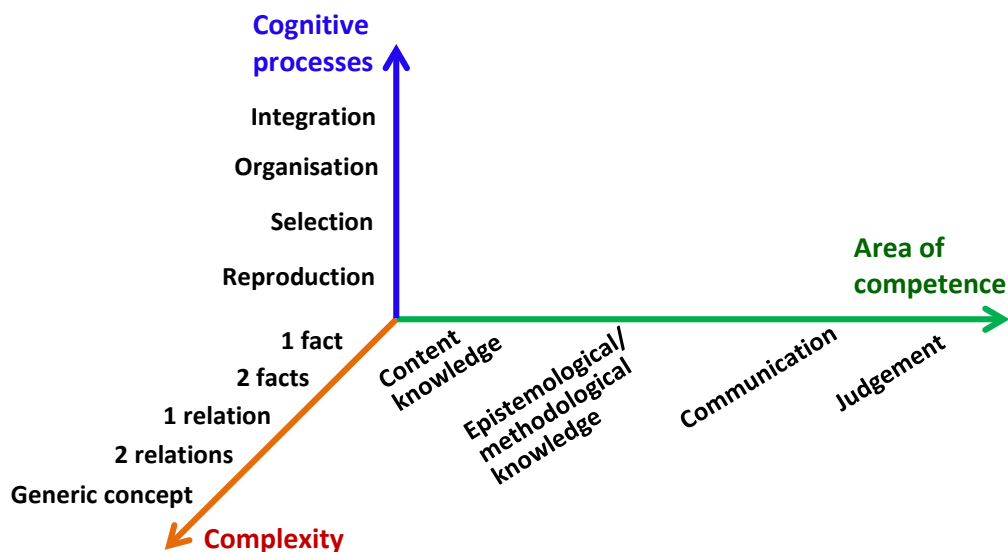

Figure 2. Competence model of ESNaS (translation by the authors)

the existing knowledge structure; e.g. proposing an experiment to test a given hypothesis would require the process of integration.

The second dimension which is supposed to be a measure for item difficulty is *complexity*. This dimension refers to assumptions about the processing of knowledge. Five levels are differentiated (in an ascending order of complexity): 1 fact, 2 facts, 1 relation, 2 relations, and generic concept.

The easiest challenge is the processing of a single fact. An item refers to the component 'fact' of this dimension if selecting or processing a single fact is sufficient to solve the item correctly. 'Relation' refers to connections between single facts, and 'concept' refers to the application of scientific concepts such as energy (Kauertz, Fischer & Siegle, 2013). It is hypothesized that it is more difficult to apply concepts than to use relations, and more difficult to process relations than make use of facts. Fundamental research in psychology indicates that this assumption is justified. Especially the model of hierarchical complexity (Commons et al., 1998) has been researched extensively (e.g. Commons et al., 2007). Bernholt, Parchmann and Commons (2009) used it successfully to predict the item difficulty for chemistry tasks. The closely related concept of complexity of ESNaS was initially introduced by von Aufschnaiter (von Aufschnaiter & Welzel, 1997; von Aufschnaiter and von Aufschnaiter, 2003). Kauertz

(2008) modified it and applied it in a physics content knowledge test. Based on Rasch-scaling, he found that task complexity correlates with the necessary grade of content knowledge. Other studies, however, show problems with the concept of complexity as a scale of expertise. Neumann et al. (2013, p. 183) could not confirm the hypothesis that an increasingly complex knowledge base describes the development of students' conceptualization of energy. Item difficulty of their energy test did not depend on item complexity (Neumann et al., 2013, p. 178).

The concept of complexity was initially developed for the competence area of content knowledge. That, however, is only one of the four competence areas of the NES. Whether or not complexity is really useful and valid to describe an increasing item difficulty in, for example, communication is a subject of discussions. The ESNaS project uses complexity to operationalize items in all of the four competence areas. Kulgemeyer (2009) proposes for communication a different concept from psycholinguistics, which is closer, the processes of language production.

Types of test items

Following the ESnaS-model for test-development a large task pool was constructed for the evaluation of the NES. There are two types of items (c.f. Kauertz et al., 2010; Schecker & Wiesner, 2013). Both need science

---

Item "Mountain climbing"

Content Information
A compound changes its aggregate state when boiling. It changes from the liquid state into the gaseous state. It depends on air pressure and temperature when this conversion occurs.

The higher the pressure, the higher is the temperature required for a solution to boil. We speak of 'boiling temperature' and 'boiling pressure'. If the pressure decreases, the boiling temperature is reduced.

On the earth the air pressure decreases with increasing height. Mountain climbers from Germany are in a high camp at the Mount Everest at 6500 m. To cook their meal, they have to boil water.

How does the boiling temperature of the water in the high camp differ from the boiling temperature in Germany?

o The boiling temperature in Germany is higher than in the high camp.
o The boiling temperature in the high camp is higher than in Germany.
o The boiling temperatures are the same.
o We cannot know, we have to conduct an experiment to decide.

Figure 3. Sample item from the evaluation of the German NES in physics (translation by the authors). Supposedly required competence area: content knowledge.

information to solve them. Type I items require knowledge students are expected to have developed in prior instruction. Type II items; in contrast; exclusively refer to information given in the task stem itself. A sample item of type II is given in Figure 3.

This type predominates the test (Schecker & Wiesner, 2013). One of the reasons for type II items is the lack of a national curriculum. The NES do not explicitly name the actual content that is expected to be taught (e.g. Newton's laws in physics) but instead the competences that are expected to be acquired (e.g. "the students use content knowledge to solve tasks and problems"(KMK, 2005, p. 11, translation by the authors)). The science syllaby of the federal states of Germany only partially overlap, and therefore the students differ in their prior knowledge. This is a major reason why the test developers concentrated on item of type II, presenting the necessary principles or laws in the task stems – at least in the competence domain of content knowledge. This approach is of course not undisputed. Some science educators state that selecting (and processing) information from the task text itself and using it to find the answer is not what science competence should be about. Such items might just refer to a science-related reading literacy (Schecker & Wiesner, 2013). On the other hand, type II items might also be an appropriate way to address the low achieving students. Other tests like PISA are often not useful to differentiate between the lowest 15 to 20 % of the population (Labudde et al., 2009). There is evidence indicating that type A and type B items refer to the same underlying ability. Ropohl (2010) used both types of items and found them to measure the same construct. A one-dimensional Rasch model could be used to form a common scale – even if a two-dimensional scale with type I and type II items was also useful to describe the data (Ropohl, 2010, p. 85-86). As could be expected, the performance in type II items shows a higher correlation to reading ability[1] and intelligence[2] than the performance in type I items[3] (Ropohl, 2010, p. 98).

The process of test development

Several studies were conducted to provide further evidence for the ESNaS-model (Kremer et al., 2012). To name just a few, Härtig (2010) researched the curricular validity of physics tasks. The influence of prior knowledge on the performance in a competence test was examined by Ropohl (2010). Neumann (2011) researched the dimensionality of competences regarding the nature of science, a sub-area of the competence area epistemological/ methodological knowledge.

The evaluation of the standards and the test development in ESNaS go hand in hand. ESNaS considers three main steps in order to develop the test items. Firstly, textbooks were analysed to ensure that the items include the topics, which are important in science teaching (Härtig, 2010). Textbooks are often seen as crucial to identify what parts of a curriculum find the way to the actual teaching (Valverde et al., 2002). Secondly, expert teachers developed test items. Thirdly, science education researchers examined and commented on these items. Several pilot studies were conducted to provide evidence for validity and reliability of the test instrument. The largest validation study used 998 test items from all three sciences in a multi-matrix design and administered them to 6845 10th grade students from 160 schools (Kremer et al., 2012).The Rasch-scaled performance indices proved that a very broad range of personal abilities could be covered with the test (Kremer et al, 2012, p. 213). After selecting items according to their infit values and their representation of the ESNaS-model, 944 test items remained.

The final version was used in parallel to PISA 2012. First results have just been published (Pant et al., 2013). The tests were administered to 44.584 students from 1.326 schools all over Germany. The results were described on a scale comparable to the one used in PISA with the average ability set to 500 point. One year of school is supposed to correspond to a gain of 25 to 30 points. The results showed disparities between the federal states (e.g. Saxony with an average of 544 points and North Rhine-Westphalia with only 476 points, corresponding to about two years of physics teaching less). There were also disparities referring to the gender: the girls outperformed the boys in content knowledge in biology (511 versus 489 points) while the boys showed better performances in mathematics (508 versus 492 point). Results with an important impact on politics are those describing social disparities. In mathematics students from families with a higher socio-economic status outperformed students from a lower social background by 82 points, this compares to nearly three years of schooling.

The development of test instruments for the standards was conducted with large effort in Germany, especially in the context of ESNaS. The main problem most researchers see in the development process is that the standards were formulated as a quick reaction to the PISA results, without prior empirical studies or an appropriate theoretical basis. All these steps had to be taken after the standards had already been published and had started to influence science teaching. Looking at the empirical process of evaluating the standards, some researchers criticize the use of the dimensions cognitive processes and complexity, and in particular

---

[1] r=.41***, p<.001, N=364 10th grade students
[2] r=.48***, p<.001, N=1260
[3] r=.28***, p<.001, N=364 for reading and r=.36***, p<.001, N=1260 for intelligence

test items that do not necessarily need prior knowledge.
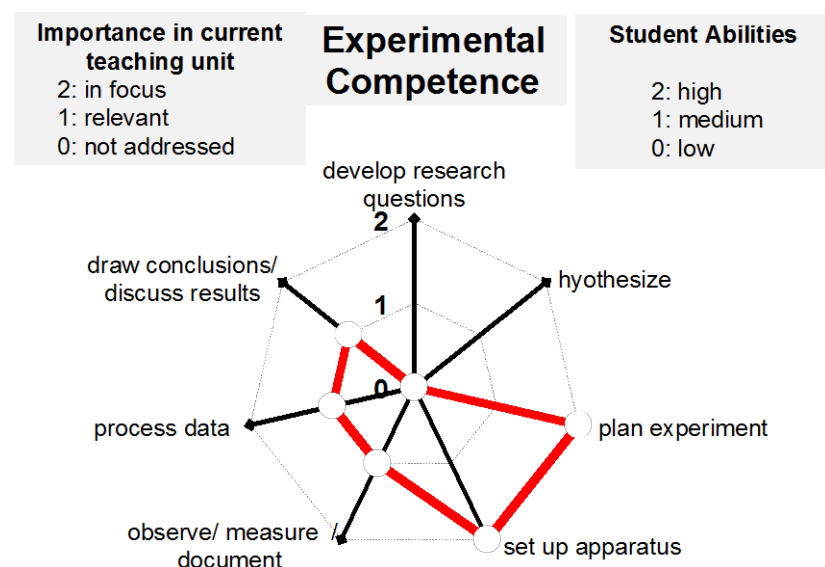
A major concern about the research in the context of

Figure 4. Model on experimental competence by Maiseyenka, Schecker & Nawrath (2013). The figure shows seven parts of experimental competence ("facets"). The numbers refer to either the importance of a certain facet of experimenting in a particular teaching unit or the level of students' abilities in the particular facet

All in all, most research in the context of the ESNaS is basic research while the development and implementation of competence-oriented teaching material is still rare. There are, however, alternative frameworks which also refer to the NES but focus more on the development of teaching materials and expand on those competence areas that have been neglected until now by the ESNaS program – especially communication.

Alternative frameworks in competence modelling in Germany

Over the last ten years since the NES were established, a number of studies have been conducted in physics, chemistry, and biology education with a focus different from large scale standard evaluation. Most of these studies researched the structure of a specific competence area in the NES, such as communication (Kulgemeyer, 2010), content knowledge (Bernholt, 2010), or judgement (Eggert & Bögeholz, 2006). Their competence models differ from the model of the ESNaS project. There are also rare examples of teaching-oriented research programs, with a specific interest in implementation and in parallel to this in the development of competence models for actual teaching. In the following we will present two of our own research programs and their result in a rough overview.

Modelling experimental competence for teaching purposes

the NES is (as stated above) that competence models are being developed for research, not for teaching purposes. The language of this research relies on the technical lingual of psychometrics. The underlying assumptions – like the concept of complexity – are rather sophisticated. Competences are graded sophistically with more than three categories for each dimension. The models are written to guide test item development. Their purpose is not to help teachers by supporting competence-oriented teaching or students' learning processes. All this might not be useful for teachers. Focussing on the implementation of competence-oriented teaching Maiseyenka, Schecker and Nawrath (2013) took a different approach. They worked together with a group of science teachers on a model of experimental competence for teaching purposes and complementary teaching material. The question whether or not the dimensions of this model can be differentiated empirically was not of their particular concern. Their model was designed explicitly to help teachers both in planning their teaching and in giving feedback to students.

Figure 4 shows the resulting model. Experimental competence in the notion of Maiseyenka, Schecker, and Nawrath (2013) and their teacher-partners can be separated into seven parts or 'facets' that can be understood as necessary skills or abilities to perform experiments in science. Starting from 'developing research questions' and following a clockwise direction in Figure 4 they represent important phases of the experimentation process. Maiseyenka, Schecker and Nawrath (2013) primarily wanted to cover the actual experimenting in school labs, not in science in general

and with a focus on actually performing experiments and evaluating data. Their model overlaps with models of scientific inquiry. Abell (2008) lists abilities students need for inquiry, e.g. asking questions, designing investigations, collecting and analysing data, using evidence to construct explanations, and communicating explanations (Abell, 2008, p. 8). Some of these abilities refer directly to the experimentation process shown in Figure 4. Experiments can be a part of the inquiry process in science, but inquiry is more than experimenting. Modelling for instance is also an important part of scientific inquiry; conducting experiments might even be the most important inquiry process in physics.

This model is supposed to be helpful for teachers to plan their teaching. Teachers who want to implement experiments in e.g. their physics lessons can look at the various facets of experimental competence and then decide which of them specifically deal with the coming lesson (cf. Figure 4). They can e.g. provide the research questions and the hypothesis so that students do not have to work them out themselves. For these two facets the teachers would note a "0" (a facet not important in this specific experiment). The focus could lie on planning an appropriate experiment to test the hypothesis: working out the experimental approach, selecting from a set of apparatus and constructing an executable experimental set-up. This focus would lead to two points for the facets 'plan experiment' and 'set up apparatus'.

One purpose of the model for teachers is to reflect on the different facets of experimental competence in their teaching. This does not mean that teachers have to regard each of the facets in every experiment they perform in their classes. But over a longer teaching period all the facets should be taken up so that students can develop competences in each of them. Teachers can also use the model as a rubric for assessment and as a structure for giving feedback to their students.

The way the model was developed differs from the way models for research purposes are developed. For the ESNaS model trained teachers just took part as item developers and had no influence on the model. The model on experimental competence started from normative considerations about experimental competence and was developed in a group of teachers and researchers ('symbiotic cooperation' c.f. Maiseyenka, Schecker, & Nawrath, 2013). Whether or not the facets can be differentiated in large-scale assessments, respectively whether or not the facets really describe different skills or abilities, was no priority issue in this project. Much more important was the question, whether the facets are useful for teachers to reflect their teaching practices and guide their tasks used in student experiments. Evaluation studies supported these functions (Maiseyenka, Schecker, & Nawrath, 2013). In

this sense the project aims at the third function of competence models stated in section 1: making educational standards more accessible for actual teaching.

Science communication competence

Communication is a competence area of the German NES, similar components can be found in the Swiss standards and also in Anglo-Saxon standards, e.g. the Australian Curriculum for Science (ACARA, 2012). Kulgemeyer (2010) developed a model of science communication competence. It is based on theoretical considerations about the process of communication in a constructivist view. The model has been validated empirically (Kulgemeyer & Schecker, 2012; Kulgemeyer & Schecker, 2013). Its central idea is the communication process shown in Figure 5. Explaining is at the core of communication competence. Explaining here means making content. In a constructivist view, good explaining makes it more likely that a scientific content is understood. It helps the addressee to construct meaning but does not necessarily lead to understanding.

A communicator – in our focus a person who wants to explain something to someone – has four variables that he or she can adapt to make a science matter comprehensible for an addressee (cf. Figure 5). The following examples are taken from Kulgemeyer & Schecker (2013). The explainer can vary the factual content aspects to be included in the explanation (e.g. the optical phenomenon of dispersion), the graphical representation form (e.g. a diagram of the ray paths), a context (e.g. a rainbow) as a situation in which the phenomenon occurs, and the code (e.g. everyday language).

If the addressee indicates that he or she could not make meaning of what the explainer said, the explainer can vary the complexity of one or more of the four variables. The explainer can e.g. switch from abstract ray diagrams to a realistic photo of dispersion or use a different example that might be closer to what the addressee already knows or is interested in. Of course that all does not necessarily lead to understanding – the explainer's efforts just make understanding more likely.

Kulgemeyer (2010) developed two assessment instruments for science communication competence: a written test and an expert-novice role play (Kulgemeyer & Schecker, 2013). The role-play puts expert students into situations where they have to explain physics phenomena to students of a younger age. For standardized testing, the novice students are coached to act in a specific way. Their task is to ask for easier or for more formal explanations, for further examples, etc. The explaining situations are video-taped and analysed with qualitative categories. The analysis focuses on the way the expert students react to the novices' questions

and prompts. Kulgemeyer and Schecker (2013) show that the reactions can be described as variations of the four variables of the communication model described above. Hypothetical criteria for good explaining, such as

What is it good for? The impact of competence modelling on science teaching in Germany

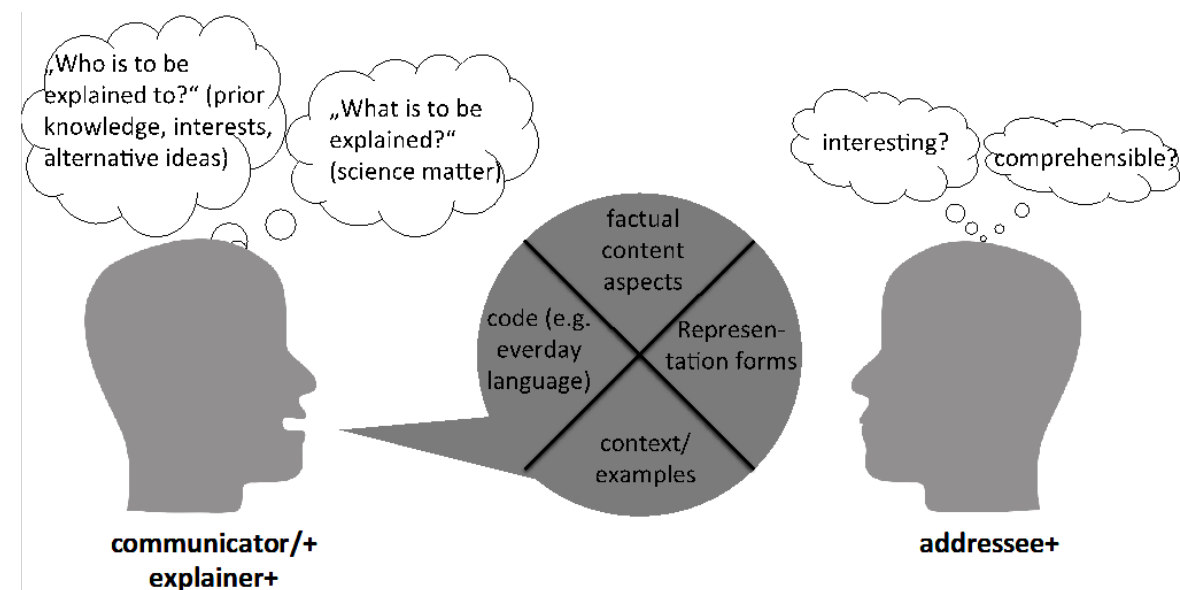In Germany educational standards are politically

Figure 5. Model of science communication process by Kulgemeyer (2010).

the use of examples, were confirmed empirically. Kulgemeyer (2010) found that individuals with a high science content knowledge only reach mid-level results in science communication competence. Explainers with just a medium level of science content knowledge were the best explainers. Further research on this surprising result is part of a follow-up research project (Kulgemeyer et al., 2012) with teacher trainees as explainers.

The path towards this science communication model follows the three steps Schecker and Parchmann (2006) describe for competence modelling. Firstly, a model was developed based on theoretical considerations. In the second step, test items were constructed to cover the model components. Thirdly, students' test performances were evaluated to check and to refine the model structure. As one result, the supposedly different components 'representations forms' and 'code' were found to depend on one another. For research purposes these two components can be integrated into a single component 'representation form', as verbal language can be treated as an especially important representation form of information among others. This reduces the number of test items required. For teaching purposes, however, it could be useful to keep the components apart, which could help teachers to develop more specific tasks for their teaching. This example shows how models for teaching purposes and models intended to be useful in research may differ from another.

CONCLUSION

seen as a means to reach a consensus about aims of education among the federal states (Schecker & Parchmann, 2007). Most German science education researchers welcomed the introduction of educational standards. The impact of competence modelling on actual science teaching has been discussed intensively and controversially (Labudde et al., 2009). In one perspective, competence modelling belongs to fundamental research about cognitive structures. Fundamental research does not have to have a direct impact on teaching. The development of competence-oriented teaching can follow, once there is sufficient empirical evidence for the structure of science competence. However, the prevailing focus on fundamental research in German science education binds many resources that are thus not available e.g. for design-based research. While there are elaborate projects on modelling competences and test development, there is a lack of projects aiming at the implementation of teaching on the basis of the educational standards.

The benefit of competence models for teaching is not undisputed either. Models for research purposes are not designed for competence-oriented teaching and it is not easy for teachers to cope with their depths of differentiation. Teachers even struggle with the competence-model published in the NES. As Hartmann-Mrochen (2011) showed, teachers hold very different perspectives, for the example the competence area 'judgement' of the NES. Participants of an in-service teacher training group did not interpret this competence area in the way the curriculum developers

or the researchers did. In post interviews after a training in competence-oriented teaching the in-service teachers still mostly thought this competence area would refer to their own judgement of student's performances. Curriculum developers and researchers had in mind the competence to make decisions in socio-scientific contexts. However, there are examples for models with a documented use for actual teaching, like the model of Maiseyenka, Schecker, and Nawrath (2013) for experimental competence. This model is the result of a symbiotic cooperation between researchers and teachers.

The first evaluation study of the German science education standards has just been published (Pant et al., 2013). There is a realistic chance that more research with a focus on implementation and development of teaching material will follow. With respect to the three purposes of competence models (Klieme et al. (2003), c.f. section 1), one, however, has to state that currently the first function, representing cognitive structures, still is the most frequent one in German research on competences.

Besides fundamental questions about the aims of science education research, another major point of discussion is a possible problem with 'teaching to the test'. When the first items were developed for evaluation of the NES, only 'content knowledge' and 'epistemological/ methodological knowledge' were included as areas of competence. Such a focus could have driven teachers to overemphasize these most familiar competence areas and to neglect 'communication' and 'judgement' in their teaching. The IQB encountered this critique by also developing items for the two remaining competence areas (Labudde et al., 2009). On the other hand, teaching to the test is not necessarily a problem if the test is valid and well-reasoned from a normative perspective in the German tradition of "Bildung" (Fischler, 2013). It remains an open question whether or not the ESNaS test fulfils these high expectations (Schecker & Wiesner, 2013) .

As always, validity is the crucial issue in test development. Whether or not written tests, like the ESNaS-test, suffice for a valid assessment in process-oriented areas like experimentation and communication is a question of research. Kulgemeyer (2010) criticizes that so far only the cognitive aspects of competence have been tested while the volitional and the motivational aspects (Weinert, 2001) have not been regarded appropriately. Even more, formulating competence models is not just a matter of quantitative research in large-scale assessments. As McLelland (1973) stated:

"Testers have got to get out of their offices where they play endless word and paper-and-pencil games and into the field where they actually analyze performance into its component parts." (McClellan, 1973, p.7)

This still remains true. There is a need for field studies to analyse and describe science competence in ecologically sound settings – and that means in schools and during actual lessons, not in studies that analyse individuals in a psychology or small groups in a science education lab. This aspect is still underrepresented in German science education research.

On the other hand, there certainly is a high potential of standardized testing. It could lead to a valid and reliable assessment of learning outcomes, which helps to formulate evidence-based recommendations for the improvement of science teaching. A broad perspective on science competences including communication and judgement enriches sciences teaching. Research results about the structure of competence in science could help curriculum developers and science educators to develop teaching materials that support specific aspects of competence more effectively. An orientation of teaching along the structure of competence might thus once be more effective than the common orientation along the structure of the scientific domain (e.g., physics).

In a nutshell, research on competence modelling in Germany so far strongly focuses on fundamental research. The full potential of the competence notion of learning science will not unfold before implementation-oriented research is also strengthened.

ACKNOWLEDGEMENT

REFERENCES

Abell, S. (2008). What students need to know and do in inquiry-based instruction. In E. Abrams, S. Southerland & P. Silva (Eds.), *Inquiry in the classroom. Realities and opportunities* (pp. 7-18). Charlotte: Information Age.

ACARA (2012). *The Australian curriculum – science.* Australian Curriculum, Assessment and Reporting Authority.

Adkinson, R. & Shiffrin, R. (1968). *Human memory: A proposed system and its control processes.* New York: Academic Press.

American Association for the Advancement of Science (AAAS) (1991). *Science for all americans.* Oxford: Oxford University Press.

Aufschnaiter, C. v. & Aufschnaiter, S. v. (2003). Theoretical framework and empirical evidence of students' cognitive processes in three dimensions of content, complexity and time. *Journal of Research in Science Teaching, 40*, 616-648.

Aufschnaiter, S. v. & Welzel, M. (1997). Wissensvermittlung durch Wissensentwicklung. Das Bremer Komplexitätsmodell zur quantitativen Beschreibung von Bedeutungsentwicklung und Lernen [Teaching knowledge by developing knowledge. The Bremen model of complexity for qualitative description of the

development of meaning and learning]. *Zeitschrift für Didaktik der Naturwissenschaften 3*, 43-58.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Eds.) (2001). *PISA 2000 - Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich [PISA 2000 – Students' basic competences in an international comparison]*. Opladen: Leske + Budrich.

Bernholt, S. (2010). *Kompetenzmodellierung in der Chemie – Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität [Modeling competences in chemistry. Theoretical and empirical considerations concerning the model of hierachical complexity]*. Berlin: Logos.

Bernholt, S., Parchmann, I. & Commons, M. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis [Modeling competences between research and classroom practise]. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 219-245.

Bransford, J., Brown, A. & Cocking, R. (Hrsg.) (2000). *How people learn: brain, mind, experience, and school*. Washington: The National Academies Press.

Chomsky, N. (1966*). Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, N. (2005). *Aspects of the theory of syntax*. Cambridge: Linguistic Inquiry.

Commons, M., Goodhearts, E., Pekker, A., Dawson, T., Draney, K. & Adams, K. (2007). Using rasch mode scaled stage scores to validate orders of hierarchical complexity of balance beam task sequence. In E. Smith Jr. & R. Smith (Eds.), *Rasch measurement: advanced and specialized Applications* (pp. 121-147). Marple Grove: JAM.

Commons, M., Trudeau, E., Stein, S., Richards, F. & Krause, S. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review, 18*, 237-278.

Dossey, J., Hartig, J., Lieme, E. & Wu, M. (2004). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003*. Paris: OECD Publications.

Einhaus, E. (2007). *Schülerkompetenzen im Bereich Wärmelehre. Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen [Student's competences in thermodynamics. Developing a test and a refining a normative model of domain-specific competences]*. Berlin: Logos.

Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz - Teilkompetenz Bewerten, Entscheiden und Reflektieren für Gestaltungsaufgaben Nachhaltiger Entwicklung [The Göttingen model of judging competence – sub-competence judging, deciding and reflecting for tasks in sustainable development] . *Zeitschrift für Didaktik der Naturwissenschaften, 12*, 177-197.

Friege, G. & Lind, G. (2003). Allgemeine und fachspezifische Problemlösekompetenz [General and domain-specific problem-soving competence]. *Zeitschrift für Didaktik der Naturwissenschaften, 9*, 63-74.

Härtig, H. (2010). *Sachstrukturen in Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests [Structures of physics textbooks as a basis for content validity]*. Berlin: Logos.

Hartmann-Mrochen, M. (2011). *Zwischen Notengebung und Urteilsfähigkeit: Einstellungen und Vorstellungen von Lehrkräften verschiedener Fachkulturen zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards (Doctoral dissertation} [Between giving grades and judging: Believes of teachers of different subjects about the competence area judging of the German national educational standards]*. Hamburg: University of Hamburg.

Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben [Factors that influence item difficulty in physics tests]*. Berlin: Logos.

Kauertz, A., Fischer, H. & Siegle, T. (2013). Erwartungen und die Wirklichkeit. Die nationale Studie zur Evaluation und Normierung der Bildungsstandards in Physik [Expectations and reality. The national study on evaluating educational standards in physics]. *Praxis der Naturwissenschaften - Physik in der Schule, 62*(5), 5-10.

Kauertz, A., Fischer, H., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I [Standards-oriented modeling of competences in science]. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 135-153.

Kauertz, A., Neumann, K. & Härtig, H. (2012). Competence in science education. In B. Fraser, K. Tobin & C. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 711-721). Dordrecht: Springer.

Klafki, W. (2006). Die bildungstheoretische Didaktik im Rahmen kritisch-konstruktiver Erziehungswissenschaft ['Bildung' and didatics in the context of critical-constructivist pedagogic]. In H. Gudjons & R. Winkel (eds.), *Didaktische Theorien* (pp. 13-35). Hamburg: Bergmann und Helbig.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. (2003). *Zur Entwicklung nationaler Bildungsstandards – Expertise [On the development of national educational standards – an expertise]*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).

KMK (Ed.) (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss [National educational standards in physics]*. München: Luchterhand.

Kremer, K., Fischer, H., Kauertz, A., Mayer, J., Sumfleth, E. & Walpuski, M. (2012). Assessment of standard-based learning outcomes in science education: perspectives from the german project ESNaS. In S. Bernholt, K. Neumann & P. Nentwig (eds.), *Making it tangible: learning outcomes in science education* (pp. 201–218). Münster: Waxmann.

Kulgemeyer, C. (2010). *Physikalische Kommunikationskompetenz. Modellierung und Diagnostik [Modeling and measuring physics commmuncation competence]*. Berlin: Logos.

Kulgemeyer, C. & Schecker, H. (2009). Kommunikationskompetenz in der Physik: Zur Entwicklung eines domänenspezifischen Kompetenzbegriffs [Communcation competence in physics. On the development of a domain-specific competence concept]. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 131-153.

Kulgemeyer, C. & Schecker, H. (2012). Physikalische Kommunikationskompetenz - Empirische Validierung eines normativen Modells [Physics communication competence – Empirical validation of a normative

model]. *Zeitschrift für Didaktik der Naturwissenschaften, 18,* 29-54.

Kulgemeyer, C. & Schecker, H. (2013). Students explaining science – assessment of science communication competence. *Research in Science Education 43*(6), 2235-2256. doi: 10.1007/s11165-013-9354-1.

Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Riese, J., Schecker, H., Tomczyszyn, E. & Walzer, M. (2012). ProfiLe-P - Professionswissen in der Lehramtsausbildung Physik. Vorstellung eines Forschungsverbundes [ProfiLe-P – A research program on teachers' professional knowledge in physics teacher education]. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung.* Retrieved from http://www.phydid.de/index.php/phydid-b/article/view/380/500

Labudde, P., Duit, R., Fickermann, D., Fischer, H., Harms, U., Mikelskis, H., Schecker, H., Schroeter, B., Wellensiek, A. & Weiglhofer, H. (2009). Schwerpunkttagung Kompetenzmodelle und Bildungsstandards: Aufgaben für die naturwissenschaftsdidaktische Forschung [Competence models and educational standards: implications for research in science education]. *Zeitschrift für Didaktik der Naturwissenschaften, 15,* 343-370.

Maag-Merki, K. (2003). Überfachliche Kompetenzen: Zur Validierung eines Indikatorensystems [General competences: validating a rubric]. *Empirische Pädagogik 17,* 123-147.

Maiseyenka, V., Schecker, H. & Nawrath, D. (2013). Kompetenzorientierung des naturwissenschaftlichen Unterrichts. Symbiotische Kooperation bei der Entwicklung eines Modells experimenteller Kompetenz [Competence-orientation in science classroom. Symbiotic cooperation in developing a model on experimental competence]. *Physik und Didaktik in Schule und Hochschule, 12,* 1-17.

MCEETYA (Ministerial Council on Education, Employment, Training and Youth Affairs) (2005). *National assessment: Program, science, year 6, 2003: Technical report.* Carlton: MCEETYA.

McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist, 28,* 1-14.

National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington: The National Academies Press.

Neumann, I. (2011). *Beyond physics content knowledge - modeling competence regarding nature of scientific inquiry and nature of scientific knowledge.* Berlin: Logos.

Neumann, K., Kauertz, A. & Fischer, H. E. (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.

Neumann, K., Viering, T., Boone, W. & Fischer, H. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching 50*(2), 162-188.

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Eds.) (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I [Comparision of german federal states. Competences in maths and science].* Münster: Waxmann.

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science 340*(6130), 320-323.

Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept chemische Reaktion - Entwicklung und Analyse von Testaufgaben [Modeling student's competences in chemical reactions – development and analysis of test items].* Berlin: Logos.

Rost, J., Walter, O., Carstensen, C., Senkbeil, M. & Prenzel, M. (2005). Der nationale Naturwissenschaftstest PISA 2003 [The national science test in PISA 2003]. *Der Mathematische und Naturwissenschaftliche Unterricht, 4,* 196-204.

Shavelson, R. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2,* 43-65.

Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz [Modeling science competence]. *Zeitschrift für Didaktik der Naturwissenschaften, 12,* 45-66.

Schecker, H. & Parchmann, I. (2007). Standards and competence models: the german situation. In: D. Waddington, P. Nentwig & S. Schanze (Eds.): *Making it comparable. Standards in science education* (pp. 147-164). Münster: Waxmann.

Schecker, H. & Wiesner, H. (2013). Die Bildungsstandards Physik. Eine Zwischenbilanz nach neun Jahren [Conclusions from nine years of national educational standards in physics]. Praxis der *Naturwissenschaften Physik in der Schule 62*(5), 11-17.

Schneider, H. (2002). Beruht das Sprechenkönnen auf einem Sprachwissen? [Does the ability to speek rely on a knowledge about language?]. In S. Krämer & E. König (eds.), *Gibt es eine Sprache hinter dem Sprechen?* (pp. 129-150). Frankfurt a. M.: Suhrkamp.

Schmidt, M. (2008). *Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. Entwicklung und Erprobung eines Testinventars [Competence modeling in energy. Developing a test.].* Berlin: Logos.

Valverde, G., Bianchi, L., Wolfe, R., Schmidt, W. & Hounang, R. (2002). *According to the book.* Dordrecht; Boston; London: Kluwer.

Weinert, F. (2001). Concept of competence - a conceptual clarification. In D. Rychen & L. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe + Huber.

QCA, Department for Education and Employment (1999). *Science: The national curriculum for england: Key stages 1-4.* London: QCA.

❖ ❖ ❖