International Journal of Education in Mathematics, Science and Technology (IJEMST)

www.ijemst.com

Assessing Teaching and Assessment Competences of Biology Teacher Trainees: Lessons from Item Development

Sascha Hasse[1], Cora Joachim[2], Susanne Bögeholz [1], Marcus Hammann[2]
[1] Westfälische Wilhelms-Universität Münster
[2] Georg-August-Universität Göttingen

To cite this article:

Hasse, S., Joachim, C., Bögeholz, S., & Hammann, M. (2014). Assessing teaching and assessment competences of biology teacher trainees: Lessons from item development. *International Journal of Education in Mathematics, Science and Technology*, *2*(3), 191-205.

IJEMST International Journal of Education in Mathematics, Science and Technology
Volume 2, Number 3, July 2014, Page 191-205 ISSN: 2147-611X

Assessing Teaching and Assessment Competences of Biology Teacher Trainees: Lessons from Item Development

Sascha Hasse[1*], Cora Joachim[2], Susanne Bögeholz[1], Marcus Hammann[2]
[1] Westfälische Wilhelms-Universität Münster
[2] Georg-August-Universität Göttingen

Abstract

In Germany, science education standards for students at the end of grade nine have been in existance since 2005. Some of these standards are dedicated to scientific inquiry (e.g. experimentation). They describe which abilities learners are expected to possess at the end of grade nine. In the USA, several documents describe standards for *Teaching Inquiry* (NGSS 2013, NRC 1996/2000/2007, AAAS 1989). Presently, comparable teaching standards for science teachers are mostly lacking in Germany. Further, there are hardly any instruments that allow for the assessment of specific competences pertaining to teaching experimental lessons and assessing student competences in experimentation. Therefore, the aim of the project described in this paper is to develop assessment instruments for biology teachers who are being trained at universities as well as in in-service teacher training programs with respect to i) analyzing experimental biology lessons, ii) planning experimental biology lessons, and iii) assessing student achievements in experimental biology lessons. The article gives insights into ongoing research with respect to assessing the quality of biology teacher education. Finally, the developed measurement instruments should allow for assessing the learning preconditions of future biology teachers. The instruments offer first starting points for the development of sensitive measures for longitudinal studies to investigate university teacher education and teacher traineeship in the subject of biology.

Key words: Science education, Biology teacher trainees, Measurement instrument, Pedagogical content knowledge, Experimentation.

Introduction

The concept of competence has received increased attention in educational research in Germany. In particular, the "assessment of competencies plays a key role in optimizing educational processes and advancing educational systems" (Koeppen et al., 2008, p. 61). Also, theoretical competence models (e.g., Bybee 1997) are presently being given an empirical foundation. Though current efforts in competence modelling and assessment have focussed on student competences mainly, teacher competences have also been closely studied. Teacher competences have received even more attention after the German Federal Ministry of Education and Research launched a funding initiative dedicated to the modeling and assessment of competences in higher education in 2012 (KoKoHS; cf. Blömeke & Zlatkin-Troitschanskaia 2013).

The present paper reports on a research project (ExMo) from this funding initiative. Its main focus is the development of measuring instruments geared at testing teaching competences and assessment competences of biology teacher trainees with regard to experimentation. Three German universities are involved in this project, i.e. University of Münster, University of Göttingen and University of Bamberg. As an intended effect, the measuring instruments are expected to contribute to improving science teacher education – an international request (European Commission 2011).

Theoretical Background and Rationale

Standards for Teacher Education in Germany

In the USA, there are several documents which focus on teaching standards in general and *Inquiry Teaching standards in detail* (NGSS 2013, NRC 1996/2000/2007, AAAS 1989). In Germany, comparable teaching

* Corresponding Author: *Sascha Hasse, sascha.hasse@uni-muenster.de*

standards are mostly lacking. While standards for teacher education exist, these standards are rather general and focus mainly on interdisciplinary and pedagogic competences. Specifically, the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK 2004) has drafted a document with eleven standards for teacher education and training. These break down to aspects of Teaching, Education, Assessment and Innovation. An additional seven standards from this document pertain to biology lessons in particular. Merely one standard is devoted to Scientific Inquiry Teaching. In addition, the *Association for Subject Education* has published a framework for standards concerning the university phase of teacher training (GFD 2005). The document describes 20 standards in the following areas: Theoretical reflection of subject-matter education, subject-matter teaching, subject-specific assessment, subject-specific communication, development and evaluation of instruction and curricula. The standards also describe rather general aims such as: "Teacher trainees can describe and explain subject-specific educational concepts in a systematic way" (GFD 2005, p. 1).

Since teaching standards and assessment standards related to scientific inquiry are mostly lacking in Germany, it was necessary to specify the existing frameworks with respect to teaching scientific inquiry and assessing student achievement in scientific inquiry classes. Specifically, considerations were made concerning the question of what biology teacher trainees should be able to do (in terms of can-do statements) when they *analyze experimental biology lessons*, *plan experimental biology lessons* and *assess student achievement in experimental biology lessons*. Subsequently, test items related to these three dimensions were developed in order to build reliable and valid measures.

Teaching Experimentation in Biology Lessons

Internationally, science educators agree that scientific inquiry is central for the acquisition of scientific literacy. In addition, educational research has documented the contribution of experimental classroom experiences for the development of the leaners' scientific literacy (Abell 2007, Hofstein & Lunetta 2004, Sandoval & Reiser 2004, Chinn & Malhorta 2002, Psillos & Niedderer 2002).

Many countries have implemented teaching standards for scientific inquiry, which underlines the importance of scientific inquiry in general and of experimentation in particular (NGSS 2013, NRC 1996, AAAS 1993, Council of Ministers of Education 1996 [Canada], Department of Education 1995 [England], Ministry of Education 1993 [New Zealand], KMK 2004 [Germany]). However, learners are often unable to meet the expectations formulated in the standards (Grigg et al., 2007, Coble & Allen 2005, Bybee & Fuchs 2006, PISA 2004). Against this background, the National Research Council has argued that the learning outcomes need to be seen in the context of classroom teaching: "What students learn is greatly influenced by how they are taught" (1996, p.28).

Central ideas for effective scientific inquiry teaching are made explicit in the National Science Education Standards (NGSS 2013, NRC 1996). In Germany, the comparable documents are less detailed – as described above – and, as a consequence, they provide less guidance for teachers who intend to teach scientific inquiry in the classroom. However, scientific inquiry teaching in German schools often draws on the principles of inquiry teaching approaches that have been published internationally (cf. Hammann et al., 2008, Sandoval & Reiser 2004, Mulhall & Loughran 2003, Colburn 1997, White & Gunstone 1992). The following two examples are intended to illustrate this point.

In Germany, the national biology education standards (KMK 2004) specify that learners are expected to be able to *form hypotheses*, *plan experiments* and *analyze data*. These competences are theoretically grounded in the SDDS-Model (Scientific Discovery as Dual Search) by David Klahr (2000). Biology teachers need to be able to support students in acquiring these competences, for example by following the recommendation that instruction mirror the phases that can be observed when scientists engage in scientific inquiry. Anderson states: "It is implied that inquiry learning should reflect the nature of scientific inquiry" (2002, p. 2). This recommendation can also be found in an important document issued at the beginning of a large national project for increasing the quality of science and mathematics education in Germany (Bund-Länder Kommission 1997).

Further, scientific inquiry can be used to teach *contents* and *methods*. The dual function of scientific inquiry is clearly visible in current approaches to teaching scientific inquiry, for example when learners are expected to "develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world" (Anderson 2002, p. 2). When students engage in experiments on seed germination, for example, they can learn about the factors responsible for this phenomenon, but also about the control-of-

variable-strategy. Scientific inquiry teaching is thus marked by instructional measures that aim at a conceptual understanding as well as an understanding of the aims and methods of scientific inquiry.

Future biology teachers should be trained to take these exemplary ideas and distinctions into consideration when planning and analyzing experimental biology lessons. These ideas and distinctions are also central for developing a measurement instrument that aims at testing teacher trainees' competences, as the two following examples show:

- In a test item concerned with assessing the competence of planning experimental biology lessons, a work sheet is depicted that a teacher wants to use in class. In the work sheet, the phase of hypothesis formation is not taken into account. Thus, the work sheet it is not systematically oriented towards the stages of scientific inquiry. The teacher trainees are asked to modify the work sheet in a way that it also promotes hypothesis formation.
- In a test item concerned with assessing the competence of analyzing experimental biology lessons, a situation is depicted where a group of learners records data that contradicts scientific findings. The teacher considers excluding the data of this group based on the rationale that incorrect data does not promote an adequate understanding of a biological phenomenon. The teacher trainees are asked to decide whether or not the teacher's intended action is appropriate. The teacher trainees are expected to recognize that it is not content knowledge alone that can be gained from an experiment. Disconfirming data can also be used to train students how to analyze data appropriately.

Item development very soon made it clear that there are multiple alternative ways to proceed when doing scientific inquiry and that it is impossible to expect teacher trainees to describe *the one and only* correct way. Item development, as indicated above, built on the idea that there are more or less effective ways of teaching scientific inquiry – and that mismatches between educational goals and procedures must be avoided, but this does not mean "that all teachers should pursue a single approach to teaching science" (Anderson 2002, p.2).

Definition of Competences

In this paper, the focus lies on teachers' *competences*, e.g., analyzing experimental biology lessons, planning experimental biology lessons and assessing student achievement in experimental biology lessons. Drawing on Weinert (2001), Klieme & Leutner (2006) and Koeppen et al. (2008), competences are defined as "context-specific cognitive dispositions that are acquired and needed to successfully cope with certain situations or tasks in specific domains" (Koeppen et al., 2008, 62).

Specifically, the competence to analyse lessons is defined as the cognitive disposition to "appropriately apprehend and assess the quality of observed lessons with regard to effectiveness" (Plöger & Scholl 2014).

Further, the competence to plan lessons is defined as the cognitive disposition to "anticipate goal-oriented actions in future situations. It is connected to the determination of prerequisites for successful actions (e.g., learning preconditions of students or the availability of materials, media, tasks) and to the thinking through of different opportunities for action in order to decide on a certain course of action" (Kiper 2012).

Finally, the *competence of assessing student achievement* is considered as the cognitive disposition to "continuously assess the level of knowledge, learning progress and performance difficulties of individual learners as well as the difficulties of different learning tasks" (Weinert 2000, p.14).

Target Group

The study described in this paper aims at assessing the competences of university students intending to become biology teachers. Future biology teachers decide at the beginning of their university studies, which teaching certificate they aim for: (i.e., high school, comprehensive school, vocational school and academic high school.) All types of biology teachers were included. Also, the sample included students from the two phases of university education (BA and MA). Several German universities from the Länder of North Rhine-Westphalia, Lower Saxony, Mecklenburg-Hither Pomerania and Bavaria participated in the pre-piloting and piloting of the measurement instruments.

Considerations for the Development of the Measurement Instruments

Connection to current research

Pedagogical Content Knowledge and Competence: In the USA and in many countries world-wide, teachers' expertise is currently being researched within the framework of Pedagogical Context Knowledge (PCK). A European contribution to PCK research is its emphasis on teachers' competences –rather than teachers' knowledge – a difference that will be further elaborated in the following part of the paper.

American researchers assume a *knowledge base of teaching* (Shulman 1986, 1987), which consists of several *categories of knowledge*, including Pedagogical Content Knowledge. The dimensions of PCK are framed differently depending on the research group. Shulman (1986, 1987), for example, names seven categories of PCK relevant for science teaching, Magnusson et al. (1999) five. The term *knowledge* seems to be the focal point of American research.

German research regarding teachers' professional knowledge utilizes the framework of international PCK research, but focusses on assessing competence. The terms *knowledge* and *competence* refer to different constructs. The term competence is defined as the "mental conditions necessary for cognitive, social and vocational achievement" (Weinert 1999, p. 26). Thus, the emphasis lies on coping with real-world problems. As a consequence, competence research focuses on problem solving skills, i.e., "all those skills required to evaluate the relevant features of a problem, so that suitable solution strategies can be selected and used" (Weinert 1999, p. 8). Without PCK however an instructor cannot be competent. "Knowledge is the necessary foundation of competence" (Weinert 1999, p.5).

PCK-models, hence, are not identical with competence models. Rather, competence models focus on a defined psychological construct (see above) and they specify the structures of a competence (structure models), levels of competence (stage models) and changes in competence through instruction and in time (development models) (cf. Koeppen et al., 2008). Structural similarities, however, can be seen, when the components / categories of PCK models are compared to the structure model of teacher trainee competences presented in this paper (i.e., analyzing experimental lessons, planning experimental lessons and assessing student achievement in experimental lessons). Specifically, it is possible to draw on the PCK-model by Magnusson et al. (1999) in order to illustrate similarities. In Magnusson's model, five components of PCK are described: *Orientation to Teaching Science, Knowledge of Science Curricula, Knowledge of Assessment of Scientific Literacy, Knowledge of Instructional Strategies* and *Knowledge of Students´ Understanding of Science*. The competences of analyzing and planning experimental lessons can be attributed to the PCK-components of *Knowledge of Students´ Understanding of Science* and *Knowledge of Instructional Strategies*. Further, the competence of assessing student achievement in experimental lessons can be related to the PCK component of *Knowledge of Assessment of Scientific Literacy*.

Projects with related Objectives: Test instruments for assessing the competences of planning and analyzing lessons focusing on scientific inquiry are rare. Prior to this project, however, it was possible to find related studies with similar research questions.

The project *Pedagogy of Science Inquiry Teaching Test* (POSITT, Cobern et al., 2014) is concerned with assessing pedagogical content knowledge of inquiry science teaching. The POSIT-Test is an important reference point for the present study, as item development for POSITT showed that it is possible to use realistic vignettes with questions related to them for a paper-and-pencil test. A similar approach to item development is presented in this paper. POSITT, however, focuses on teacher trainees' preferences regarding different teaching strategies and, assesses so-called *teachers´ orientations*. In ExMo, in contrast, realistic teaching vignettes are used in order to assess teachers´ competences.

In the project *Professional Minds,* Oser (2010) examines the quality of complex competence profiles (not individual competences) of teachers, which include cognitive aspects (e.g. *clarity of task*) as well as affective aspects (e.g. *acceptance, empathy*). ExMo, in contrast focusses on individual competences which are defined as cognitive dispositions.

Teachers' analyzing competence is currently being investigated in a project by Plöger and Scholl (2014). This study, however, is not concerned with a specific, subject-specific procedural competence like experimentation. Instead more universal aspects related to analyzing classroom situations are being examined. Plöger & Scholl (2014) use the model of *hierarchical complexity* (Commons 2008), and distinguish between *horizontal*

complexity and *vertical complexity. The same framework* is also used in the study presented here for developing items and for coding the answers (see *options for the coding of open tasks*, p.7).

Seidel et al. (2011) investigate teachers' perception of classroom situations. Specifically, classroom situations are presented in the form of video vignettes and teachers are asked to analyze them. In this study, rather general criteria (as opposed to subject-matter specific criteria) are used, such as e.g. the difference between describing and explaining a classroom situation. The distinction, however, is well taken. It is relevant for item development in the study presented here. The concept of professional perception (Goodwin, 1994; Sherin, 2002) states that the mere description of a lesson puts lower requirements on a teacher than explaining and predicting. This aspect of analysing a lesson is taken into account in ExMo for the development of tasks and code manuals as well.

Baer et al. (2011) investigate teacher trainees' knowledge about important aspects of planning a lesson. The focus of their research is the teacher trainee's knowledge of important concepts (de Jong & Ferguson-Hessler, 1996), for example knowledge of teaching methods and curricula. The level of specificity, however, required for answering the items, is very general. The teacher trainees, for example, can solve an item by simply stating that it is important to plan longer teaching units (as opposed to individual lessons) and that it is important to make choices against the background of their knowledge of curricula. Also, subject-matter specific aspects regarding experimentation are not taken into account in this project.

Dübbelde (2013) examines diagnostic competences of biology teacher trainees concerning the domain of knowledge acquisition. The project aims at developing a test instrument with closed task types for status and process diagnostic competences. Among other things it is recorded how far biology teacher trainees assess students' results and work processes when experimenting with the help of given evaluation criteria. Dübbelde pursues a partly similar aim as we do within ExMo regarding assessment competences. In her test instrument teacher trainees are given, for instance, a worksheet filled out by two students to document the steps of their experiment. The teacher trainees are asked to assess the students' results with regard to the given criteria. For each criterion, the teacher trainees have to choose one of three (or four) alternative answers. For instance, they have to assess whether the students' hypothesis is related to the research question by ticking off "*yes"*, "*no"* or "*don't know"*.

The test instrument used in Dübbelde (2013) includes comparable criteria pertaining to experimentation as the ExMo test instrument. In ExMo, however, it is of central interest to find out to what extent the teacher trainees know (and activate on their own) criteria with respect to experimentation, typical preconceptions and difficulties students have when experimenting. In addition, we are interested in knowing to what extent teacher students are able to independently utilize these for the assessment of students' achievements. For a differentiated evaluation of the teacher trainees' assessment cognitions open tasks are used in ExMo. The tasks describe students´ performance in experimenting and then ask the teacher trainees to assess either the formation of hypotheses, planning of experiments or data analysis. For this, the teacher trainees have to be aware of the criteria and apply them correctly and in a sophisticated manner.

Selection of Subject-Specific Content

The teaching vignettes focus on biological topics that can be found in the curricula of most Länder in Germany. Also the biological topics chosen can be combined with experiments pertinent to students. For the grades 5-6, seed germination was chosen, for grades 7-8 photosynthesis and for grades 9-10 enzymes.

Central Challenges

First attempts at item development quickly showed two major challenges, which deserve closer study:

1) In order to assess the teacher trainees' competence to analyze lessons, the complexity of the situation has to be reduced to some degree so that it is possible to code the answers of the teacher trainees´ test. At the same time, the complexity shouldn't be reduced too far so that the realistic character of classroom situation doesn't get lost. The aim is to assess a person's competence to solve real-world problems *and* arrive at answers that can be coded.

2) In order to assess the competence of planning lessons, the openness of planning decisions must be restricted to some degree in order to arrive at answers that can be coded. However, the situation should

not be reduced too much, so that the character of the situation still classifies as real-world problem solving. This situation is analogous to the situation described under challenge 1.

Dealing with Challenge #1

In order to sufficiently reduce the complexity of analyzing classroom situations, the decision was made to explicitly state which competence the teacher in the teaching vignette intends to promote when teaching an experimental lesson. Also, the question that needed to be answered was framed in a way that decreased the possibility of variation. In a current item (see appendix: *Task 1*), the description can be found that a teacher has three different options in order to promote the student competence of planning experiments independently. The teacher trainee's task is to judge which approach is the most suitable and give reasons for their decision.

During item development two further insights were gained: Multiple choice questions proved unsuitable because it was found possible to answer them through logical reasoning and reading skills alone (see appendix: *Task 2*). Also, open-answer tasks, which did not specify the competence the teacher intends to promote, allowed for too much variation in answers so that coding the answers proved impossible.

Dealing with Challenge 2

Similar to challenge 1, it was necessary to find a way of limiting the variation in possible answers. In particular, the item contains a description of an experimental lesson. The teacher trainees are encouraged to plan alternatives or suggest changes because specific aspects of the plan contain flaws or mismatches between intended aims and specific aspects of the lesson. Generally, items assessing the competence to plan experimental lessons, also state which experimental competence the teacher intends to promote.

Options of Coding Open Tasks

When coding the answers we utilized Commons' (2008) concept of complexity. Commons describes that it is possible to distinguish complexity in two ways: *Horizontal complexity* implies that several pieces of information are processed on the same level, while *vertical complexity* entails a processing of information on different levels. With regard to teaching and assessing competences of teachers, this model of complexity can be applied as follows: When analyzing, planning and assessing, teachers must constantly take several unrelated aspects into account. This may entail e.g. aspects related to subject matter, social aspects and methodological teaching aspects. A teacher has to consider several students´ conceptions that are independent from each other or diagnose student errors, which occur simultaneously but independent from each other (=horizontal complexity). The more aspects there are that need to be considered, the greater is the challenge for the teacher. It is not only the amount of tasks to be managed simultaneously but also the difficulty of an individual task, which influences the complexity of the challenge. Thus it is easier e.g. to simply name an occurring problem rather than give a well-founded explanation of the causes of the problem (=vertical complexity).

An exemplification of the coding manual of a task that encompasses both horizontal and vertical complexity can be found in the appendix (see *Task 1).*

At the end of the task, teacher trainees are required to rank three options from the easiest to the most difficult and to describe which aspects of planning an experiment are responsible for the different levels of difficulty. The following three aspects can be distinguished for differentiating between the difficulty of the three options (cf. Hammann et al., 2007):

1. Does the teacher tell the students which factors need to be examined [easier] or do the students have to determine the factors themselves [harder]?
2. Do the students have to examine one factor [easier] or a several factors [harder]?
3. Do the students have to plan a small number [easier] or a large number [harder] of experimental setups?

The maximum score for this task is 4 points. Mentioning the three difficulty-generating aspects (horizontal complexity) and giving reasons for the three difficulty-generating aspects (vertical complexity) are scored with one point each, as is the correct ranking of the three aspects. The assumption underlying this coding is that on the one hand a teacher needs well-founded theoretical knowledge about the difficulty-generating aspects of experiment planning while on the other hand especially the performance during the lesson is key for students'

learning success. Hence, a teacher trainee who names the correct and consequently sensible order for the practical application during a lesson, but only names two of the difficulty-generating aspects receives the same number of points as a teacher trainee who names all three aspects but does not arrange the options in an appropriate way.

The coding guide provides guidelines as well as anchor examples and contrasting examples, as specified by Bühner (2011).

Item Development

Iterative Process

According to Wilson (2005), item development is an cyclical process with four "*building blocks"* (i.e., *construct maps, item design, outcome space and measurement model*). The results of each step in the process inform the next step. Also, the process is iterative and the cycle may be repeated multiple times.

Item Development for nine Facets of Teaching and Assessment Competence

Nine facets (see *Table 1*) arise as a result of crossing three teachers' competences with three student competences. Prior to item development, a framework for item development was drafted in order to provide a systematic basis that was meant to ensure the subsequent comparability of all tasks in data analysis (Murphy & Davidshofer, 2005; Gruijter 2008). This framework states, for example, that the item development follows the approach of rational item construction (Kline, 2005), that items require open-responses, and that items start with a description of a realistic situation.

Table 1: Facets of teaching experimentation in biology

Teachers competences / *Students competences*	*Analyzing experimental lessons* Analyzing teachers´ decisions that aim at ...	*Planning experimental lessons* Planning instructions that aim at ...	*Assessing students achievements in experimental lessons* Assessing the quality of ...
Forming hypotheses	...teaching students how to form hypotheses	...teaching students how to form hypotheses	...hypotheses formed by students
Planning experiments	...teaching students how to plan experiments	...teaching students how to plan experiments	...experiments planned by students
Analyzing data	...teaching students how to analyze data	...teaching students how to analyze data	...students interpretations gained by analyzing data

Formulation of concrete Requirements for Teaching Experimentation

In general, the development of a test for assessing complex features must always be preceded by a specification of the object of measurement (cf. Kline, 2005). Taking into consideration the relevant specialized literature (e.g., Carey et al., 1989; White & Gunstone 1992; Gott & Duggan, 1995; Driver el al., 1996; Colburn, 1997, Chen & Klahr, 1999, Kanari & Millar, 2004, Bybee et al., 2006, Hammann et al., 2008; Ford, 2008; Gyllenpalm et al., 2010), central requirements for biology teachers when teaching experimentation were organized with regard to the nine facets.

The latter shall be illustrated by means of an example for the facet of *Analyzing teachers´ decisions that aim at teaching students how to form hypotheses*: A biology instructor should be able to...

- …evaluate and analyze the challenges in planning different experimental courses of action.
- …identify the aspects that constitute the range of complexity of different tasks. This especially includes the number of variables to be tested and the number of experimental setups to be compared as well as naming of the variables to be tested.

This concrete requirement was operationalized in the test item discussed above (see appendix: *Task 3*).

Discussing Prototypical Tasks with Experts

Following the development of prototypical items, a multi-day workshop was conducted. During this workshop a framework for the item development and prototypical items were introduced and discussed. As part of this meeting all prototypical tasks were discussed, modified or excluded, if they proved unsuitable for the assessment of the targeted competence.

In addition, the tasks and items for the evaluation of assessment competence were tested in an expert panel for the validity of their content. Six experts (among them three scientists and three teachers) came to the conclusion that the lesson vignettes can be considered realistic and the tasks may be considered part of the interesting collectivity of possible tasks for assessment competence regarding experimentation. The results of this expert survey were taken into account in the further development of items.

Studies of Thinking-Aloud Protocols

The aim of think aloud protocols (cf. Ericsson & Simon, 1980 & 1999) is to assess people's' cognitive processes (Hussy et al., 2010), for example in order to make sure that the items are suited to initiate the processes that are expected to occur when analyzing a lesson, planning a lesson and assessing learning outcomes.

In the study, 32 biology teacher students (16 people worked on items concerned with analyzing and planning at the University of Münster, an additional 16 people worked on items concerned with assessment at the University of Göttingen) we presented with 8 or 10 items each. All test persons took part in the study individually and received a standardized methodological instruction to the study of thinking aloud in the beginning. The think aloud protocols were analyzed qualitatively in order to refine items for the following quantitative studies.

Item Piloting and Analysis

Sample and Goals

The piloting of the developed tasks encompassed 2 subsequent studies: In the pre-pilot, 51 students of the Universities of Münster and Göttingen participated. In total 60 items were tested. Each teacher trainee received a test booklet with 9 items that either required analyzing and planning experimental lessons (N=27) or assessing student achievement in experimental lessons (N=24). The aims of the pre-pilot were the advancement of the scoring guides and the optimization of tasks.

In the second study, the pilot study, 160 students from six German universities have participated so far. In this phase, each testing booklet contains 9 items concerning analysis and planning or the assessment of students' achievements.

Work so far

The project ExMo currently moved on to its second pilot stage. The completion of assessment and a thorough analysis of the data, which allows for analyses of reliability and validity of the testing instrument, are still pending. When the data is available, a comparison between Bachelor and Master's students will be conducted in order to investigate whether Bachelor students have less developed competences than Master students. Should this be the case it will be considered indicative of acquirable cognitive competences having been measured rather than intelligence. The preliminary results of the study with thinking aloud indicate that the competences increase over the courses of university education and that students acquiring a teaching degree for academic high school perform better than students acquiring a teaching degree for any other school type. These findings are descriptive and explorative and they were not statistically tested.

Conclusion

Requirements for the development of paper-and-pencil tasks were described with respect to the assessment of teaching competences (analysis and planning of lessons). Specifically, lessons from item development showed that it is necessary to restrict the openness of the planning situation to a degree where it is possible to code whether the planning decision was made on the basis of subject-matter specific knowledge. Furthermore, it is

necessary to specify the learning objectives when assessing analyzing competence so far to allow judgment on whether or not the given classroom scenarios were appropriately analyzed.

The approach seems promising despite it being impossible to report on inter-rater agreement, reliability and validity at this point. It is presumably possible to transfer the principles of item development to other subject-educational contexts, e.g. the planning and analysis of non-experimental biology lessons. Science educators, who are interested in the measuring of competences, are encouraged to test this approach and apply it to other areas.

References

AAAS. (1993). *Benchmarks for science literacy*. Washington D.C.: American Association for the Advancement of Science.

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell, & N. G. Lederman (Eds.), *Research on science teacher education* (pp. 1105-1149). New York: Routledge.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13*(1), 1-12.

Baer, M., Kocher, M., Wyss, C., Guldimann, T., Larcher, S., & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg. *Zeitschrift für Erziehungswissenschaft, 14*(1), 85-117.

Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.) (2013). *The German funding initiative "Modeling and Measuring Competencies in Higher Education": 23 research projects on engineering, economics and social sciences, education and generic skills of higher education students.* (KoKoHs Working Papers, 3). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (1997). Gutachten zur Vorbereitung des Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts". Bonn.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte Auflage). München: Pearson Studium.

Bybee, R. W., Taylor, J. A., Gardner, A., van Scotter, P., Carlson Powell, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness.* Unpublished manuscript.

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching, 43*(4), 349-352.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of grade 7 students´ understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*(special issue), 514-529.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098-1120.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175-218.

Coble, C. R., & Allen, M. (2005). *Keeping America competitive: Five strategies to improve mathematics and science education*. Denver: Education Commission of the States.

Cobern, W. W., Schuster, D. G., Adams, B., Skjold, B., Mugaloglu, E. Z., Bentz, A., & Sparks, K. (2014). Pedagogy of Science Teaching Tests: Formative Assessments of Science Teaching Orientations. *International Journal of Science Education*. http://bit.ly/RE95xZ

Colburn, A. (1997). How to make lab activities more open ended. *CSTA Journal,* (Fall 1997), 4-6.

Commons, M. L. (2008). Introduction to the model of hierarchical complexity and its relationship to postformal action. *World Futures, 64*(5-7), 305-320.

Council of Ministers of Education. (1996). *Common framework of science learning outcomes K-12 (draft).* Victoria B.C.: Ministry of Education, Skills and Training.

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

Department of Education. (1995). *Science in the national curriculum*. London.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham u.a.: Open Univ. Press.

Dübbelde, G. (2013). *Diagnostische Kompetenzen angehender Biologie-Lehrkräfte im Bereich der naturwissenschaftlichen Erkenntnisgewinnung*. http://nbn-resolving.de/urn:nbn:de:hebis:34-2013122044701.

Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis: Verbal reports as data* (Rev, 3 print ed.). Cambridge, Mass. u.a.: MIT Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.

European Commission (2011). In Eurydice (Ed.), *Naturwissenschaftlicher Unterricht in Europa*. Brüssel: Exekutivagentur Bildung, Audiovisuelles und Kultur.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education, 93*(3), 404-423.

Gesellschaft für Fachdidaktik e.V. (2005). *Fachdidaktische Kompetenzbereiche, Kompetenzen und Standards für die 1. Phase der Lehrerbildung (BA+MA)*

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.

Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum*. Open University Press Buckingham.

Grigg, W., Donahue, P., & Dion, G. (2007). In US Department of Education (Ed.), *The nation's report card: 12th-grade reading and mathematics*, 2005. NCES 2007-468. Washington D.C.: ERIC.

Gruijter, Dato N. M. de, & Kamp, L. J. T. v. d. (2008). *Statistical test theory for the behavioral sciences*. Boca Raton u.a.]: Chapman & Hall/CRC, Boca Raton u.a.].

Gyllenpalm, J., Wickman, P., & Holmgren, S. (2010). Teachers´ language on scientific inquiry: Methods of teaching or methods of inquiry. *International Journal of Science Education, 32*(9), 1151-1172.

Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education, 42*(2), 66-72.

Hammann, M., Ganser, M., & Haupt, M. (2007). *Experimentieren können. kompetenzentwicklungsmodelle und ihre nutzung im unterricht*. Geographie Heute, (255/256), 88-91.

Hammann, M., Phan, T. T. H., Ehmer, M., & Bayrhuber, H. (2006). Fehlerfrei experimentieren. *Mathematischer und Naturwissenschaftlicher Unterricht, 59*(5), 292-299.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28-54.

Hussy, W., Schreier, M., & Echterhoff, G. (2010). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. Berlin: Springer.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748-769.

Kiper, H. (2012). *Unterricht planen, durchführen, auswerten - Überlegungen zur lernwirksamen Unterrichts planung*. In K. Bauer, & N. Logemann (Eds.), (pp. 151-182). Münster: Waxmann Verlag.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge: MIT Press.

Klieme, E., & Leutner, D. (2006). *Kompetenzmodelle zur erfassung individueller lernergebnisse und zur bilanzierung von bildungsprozessen. beschreibung eines neu eingrichteten schwerpunktprogramms der DFG* [competence models for assessing individual learning outcomes and evaluating educational processes. description of a new priority program of the German research foundation, DFG]. Zeitschrift Für Pädagogik, 52(6), 876-903.

Kline, T. (2005). *Psychological testing*. Thousand Oaks: SAGE.

KMK. (2004). Bildungsstandards im Fach Biologie für den mittleren Schulabschluss. Beschluss vom 16.12.2004 Luchterhand.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. Zeitschrift Für Psychologie/*Journal of Psychology, 216*(2), 61-73.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95-132). Dordrecht: Kluwer Academic Publishers.

Ministry of Education. (1993). *Science in the New Zealand curriculum*. Wellington, New Zealand: New Media.

Mulhall, P., Berry, A., & Loughran, J. (2003). Framework for representing science teachers´ pedagogical content knowledge. *Asia-Pacific Forum on Science Learning and Teaching, 4*(2)

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6, international ed.). Upper Saddle River, NJ: Pearson Education International, Prentice-Hall.

National Research Council. (2000). *Inquiry and the national science education standards. A guide for teaching and learning*. Washington D.C.: National Academy Press.

National Research Council. (2007). In Duschl R. A., Schweingruber H. A. and Shouse A. W.(Eds.), *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: National Academies Press.

National Research Council. (1996). *National science education standards*. Washington D.C.: National Academy Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington D.C.: The National Academy Press.

Oser, F., Heinzer, S., & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft, 38*(1), 5-28.

PISA-Konsortium Deutschland. (2004). PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland– Ergebnisse des zweiten internationalen Vergleichs.

Plöger, W., & Scholl, D. (2014). Analysekompetenz von Lehrpersonen – Modellierung und Messung. *Zeitschrift für Erziehungswissenschaft, 17*(1), 85-112.

Psillos, D., & Niedderer, H. (2002). *Teaching and learning in the science laboratory*. Dordrecht: Kluwer Academic Publishers.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*(3), 345-372.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119-150.

Shulman, L. S. (1986). Those who understand: Knowledge growths in teaching. *Educational Researcher, 15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Seidel, T, Stürmer, K, Blomberg, G, Kobarg, M, Schwindt, K. (2011). "Teacher learning from analysis of videotaped classroom situations: does it make a difference whether teachers observe their own teaching or that or others?" *Teaching and Teacher Education, 27*: 259-267.

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). Seattle: Hogrefe & Huber Publishers.

Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz 2*, 1-16

Weinert, F. E. (1999). Konzepte der Kompetenz. Gutachten zum OECD-Projekt "Definition and Selection of Competencies: Theoretical and conceptual foundations (DeSeCo)".

White, R., & Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. London: Routledge.

Appendix

Task 1: Item to assess the competence Analyzing experimental lessons related to the teaching objective planning experiments (current version)

Analysing experimental lessons Planning experiments

Mr. Hahn teaches biology in a sixth grade. He started working on seed germination and would like his students to experiment actively in class. Mr. Hahn wants to focus on the promotion of the competence to **plan experiments**.

As the learning group does not have many experiences in experimenting independantly, he compares three different possible approaches to teach students how to plan experiments (see A-C).

He considers what the students are required to do in each of the three options. He would like to rank them from the easiest to the hardest concerning the difficulty of planning experiments.

Three approaches to initiate the planning of experiments

A) The teacher hands out bean seeds to the students. The learners are instructed to find out by which factors the seed germination is affected.

B) The teacher shows the students a flower pot with soil containing beans that has germinated and describes the precise conditions of the germination. He hands out bean seeds to the students. The students are instructed o find out whether or not the soil is required for seed germination.

C) The teacher hands out bean seeds to the students. The students are instructed to find out whether or not seeds require light and warmth for germination. In addition the students are instructed how to handle with factors other than light and warmth.

Task:

Rank the three options from the easiest to the hardest concerning the difficulty of planning experiments.

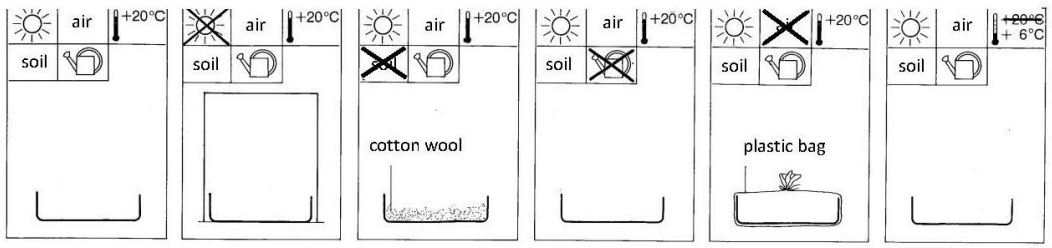
Analyze which aspects determine the level of difficulty of the three approaches to teaching students how to plan experiments!

Task 2: Multiple choice – Item to assess the competence *analyzing experimental lessons*

A teacher wants to improve the students´ competencies of carrying out biological experiments by using the worksheet depicted below.

Essential factors for the germination of bean seeds:

1. Put 10 bean seeds into each petri dish and keep it under the stated conditions
2. Make observations every second day about the germination ("+" for each germinated seed "–" for each not germinated seed)
3. Write down the conclusions of your experiment and complete the answer sentences



Datum

For the germination of bean seeds

These factors are essential: _____

These factors are NOT essential:

Task:

Which of the following competencies will be fostered by answering the three tasks on the worksheet?

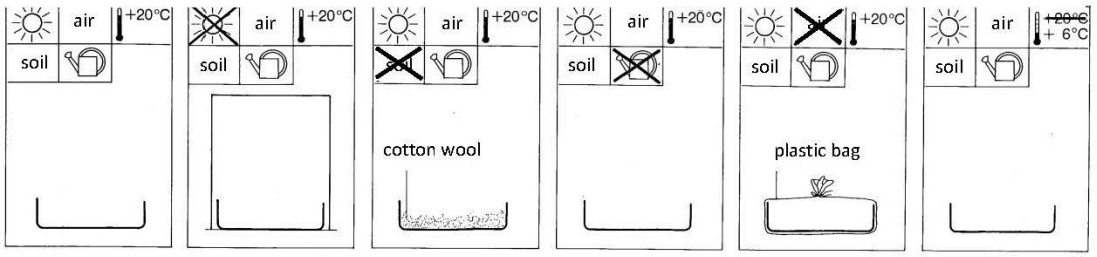
	Check the correct answers
Asking scientific questions	( )
Forming hypotheses	( )
Planning experiments	( )
Keeping precise records	( )
Analyzing data	( )
Analyzing experiments critically	( )
Relating results to new questions	( )

Task 3: Open format item to assess the competence *analyzing experimental lessons*

A teacher wants to improve the students´ competencies of carrying out biological experiments by using the worksheet depicted below.

Essential factors for the germination of bean seeds:

1. Put 10 bean seeds into each petri dish and keep it under the stated conditions
2. Make observations every second day about the germination ("+" for each germinated seed "–" for each not germinated seed)
3. Write down the conclusions of your experiment and complete the answer sentences



Datum

For the germination of bean seeds

These factors are essential: _____

These factors are NOT essential:

Task:

Are the three tasks on the worksheet suitable to improve the pupil's competencies of conducting experiments? Give reasons for your assertions!