

Informal Assessment of Competences in the Context of Science Standards in Austria

Iris Schiffl

School of Education, University of Salzburg, Austria

Copyright©2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Science standards have been a topic in educational research in Austria for about ten years now. Starting in 2005, competency structure models have been developed for junior and senior classes of different school types. After evaluating these models, prototypic tasks were created to point out the meaning of the models to teachers. At the moment, instruments for informal competency diagnosis are developed. The term “informal competency diagnosis” is used to distinguish this kind of diagnosis, which is carried out by the teachers themselves, from nationwide formal competency tests. One of these instruments for informal diagnosis is the IKM (instrument for informal competency measurement). It is developed for the informal diagnosis of science competences in junior classes. This article deals with the question if the underlying construct of the IKM can be supported through empirical data. Therefore the situation of science standards in Austria is described first to illustrate the context in which the development of the IKM took place. Then, the underlying theoretical construct is introduced and detailed information about the diagnosis tool is given. Later, the empirical evaluation of the theoretical construct gets depicted and discussed.

Keywords Science Standards, Competences, Informal Diagnosis, Diagnosis Instruments

1. Science Standards in Austria

Science standards have been developed in Austria since 2005. They can be seen as a reaction to the average results of Austrian students in PISA and TIMSS [1,2,3,4]. This reaction is known as “PISA-shock” and resulted from the perceived gap between the amount of money spent for the school system and the results of Austrian students in international studies. One of the reactions of the government was the introduction of standards to control the output of the school system in order to justify political decisions. Although standards testing is used in some subjects (mathematics, German and English, but not for biology, chemistry or physics), no high stakes testing takes place as the results of standards testing do not influence students’

further school career [5].

In all grades of the Austrian school system biology, chemistry and physics are taught separately, but science standards include all three subjects jointly to emphasize the aspects these subjects have in common [6]. The situation of science standards in Austria is a little bit complicated as there is no legal obligation to use standards in this area. Officially, standards are only prescribed for mathematics and German at the end of primary school and for mathematics, German and English at the end of junior classes [7,8]. This so called standards edict also states obligatory nationwide standards tests. But as science is not part of this edict, no official tests have to be conducted. Though there is no legal obligation, standards have found their way into Austrian biology, chemistry and physics classes due to other obligations in teaching, like the preparation of competency-orientated annual plans. These plans are called competency-orientated as they connect the required competences to the topics of the curriculum. Besides, in spring 2015 competency orientated A-levels, which are the final exams, before Austrian students are allowed to attend university, were introduced. Consequently, tasks have to contain the competences of the competency model for senior classes and they have to be divided into three parts: knowledge reproduction, application and reflection [9]. To prepare students for these final exams, teachers have to consider standards in their teaching, although there is no legal obligation to do so.

Standards are operationalized in competences. In Austria, competences are described according to Weinert [10] as learnable cognitive skills and abilities, which enable the students to solve problems successfully and responsibly in varying situations [8]. When developing science standards, the first step was the creation of competency structure models, first for senior classes for schools providing job-related education, then for junior classes and years later for senior classes of schools providing higher general education [11]. While the structure models for junior classes and senior classes of schools providing job-related education were evaluated empirically, no evaluation is planned for the model stated for senior classes of schools providing higher general education. The competency model for junior classes was used as underlying structure for the development of

complexity levels, which are important for the creation of diagnosis instruments, like the IKM (Instrument for informal measurement of competences), which marks the current step of development [12,13].

1.1. From Competency Structure to Competency Diagnosis Instruments

Evaluated competency structure models are the basis for stating complexity levels [14]. Complexity levels are deviated from competency structure models (requirement – see figure 1) and describe factors, which are relevant for the difficulty of competences. In Austria the competency structure model of junior classes was used as underlying structure. The competency structure model for junior classes is divided into three dimensions [15]: topics, competencies of acting and requirements (figure 1). The most important part of the model is “competencies of acting” as they contain all competencies and their structure. Competencies are arranged in three different categories: organizing knowledge, gaining information through inquiry and drawing conclusions. Each category is defined by four competency descriptions. Each description contains more than one competence (for example: “depicting information by using different kinds of representation, extracting information from different kinds of representation and communicating the extracted information”). For measurement they have to be divided (for the example mentioned above into three competences: “depicting information by using different kinds of representation”, “extracting information from different kinds of representation” and “communicating information from different kinds of representation”).

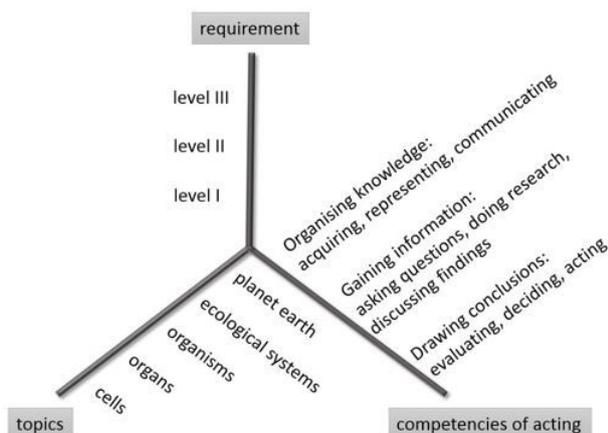


Figure 1. Competency structure model for junior classes with biological topics [16]

There are two different possibilities of stating complexity levels: First, the same quality of complexity levels can be proposed for all competencies. This possibility can be found for example in the German ESNaS Model [17,18]. Second, complexity levels can be stated for each competence individually [19,20]. When stating individual complexity levels, they can be taken a priori from literature or they can be derived from empirical data a posteriori. For the

development of the Instrument for informal competency measurement (IKM), complexity levels were phrased individually for each competence by using literature. Because of special requirements of the IKM like computerized testing and the use of special task designs (for example multiple choice), not all competences of the competency structure model could be taken into account for the use in the diagnosis instrument. The selected competences and their complexity levels are shown in table 1 [12].

The first three competences shown in table 1 are part of the “organizing knowledge” competence category, followed by five inquiry competences and two competences of the “drawing conclusions” category.

The “naming and describing natural phenomena” competences refer to the quality of speech (everyday language or science terminology) for creating complexity levels. Unfortunately, you cannot always clearly distinguish between these kinds of languages as sometimes, the same words are used in everyday language and in scientific terminology polysemously [21,22,23]. Lemke [24] advises the use of everyday language and science-related language in teaching. One characteristic of science teaching is the use of special scientific terms, which makes understanding more difficult for students [25]. The second complexity factor is the use of passive or active vocabulary [26].

Depicting information from logical representations is part of the diagram competence. The second part of this competence is called “extracting information from graphs and diagrams” [27]. Depicting information is normally operationalized in open tasks in which students have to create diagrams on their own [28,29,30]. In contrast, complexity levels of “extracting information” can easily be phrased in multiple choice items. But for now, only the first part – depicting information – was included in the IKM. Depicting information using graphs or diagrams is challenging for students especially if they have to create a graph from given data on their own [29]. The first problem students may have with this task is the choice of the right kind of graph. In a study by Baker, Corbett and Koedinger [29] conducted in 8th and 9th grade, only 25 % of the students were able to choose the right kind of representation from four given possibilities. If the students were able to choose the right kind of representation, the next difficulty would be the creation of an adequate frame for recording data. Many problems concern the correct metric of the scales or the correct naming of the axes [30,31]. After creating the frame, the next step is the entering of data points. In a study by Kerslake [31], between 89% and 95 % of thirteen to fifteen year old students were able to enter data points in a coordinate system correctly. Following these results from literature, the first complexity level for the IKM is the entering of data in a given frame, the second deals with main criteria of representation (axes, scales, naming of axes) and for level three, students have to be able to depict information independently using a correct form of representation.

Table 1. Competences used in IKM and their complexity levels

Competency	Levels
Naming natural phenomena	Naming natural phenomena by using everyday language
	Naming natural phenomena by using given science-specific terms
	Naming natural phenomena by using science-specific terminology
Describing natural phenomena	Describing natural phenomena by using everyday language
	Describing natural phenomena by using science-specific terminology
	Describing natural phenomena by using underlying concepts
Depicting information by using different kinds of logical representations	Depicting information within given representational structures
	Acknowledging main criteria of a representation and depicting information
	Depicting information by using self-selected representations
Formulating hypotheses	Formulating hypotheses from a given issue
	Formulating hypotheses from a given issue with respect to influencing factors
	Formulating hypotheses from a given experiment with respect to influencing factors and their variation
Planning an experiment	Reproducing the plan of a given experiment
	Planning an experiment with given instruments for measurement
	Planning an experiment independently
Performing measurements	Reading data from a given measuring instrument
	Choosing a suitable measuring instrument and performing measurements
	Choosing a suitable measuring instrument and performing measurements, considering limits of measurement and switching between measurement units.
Writing records	Filling in the most important points of a given experimental sequence in a given experimental report
	Filling in the most important points of a given experimental sequence in an experimental report independently.
	Filling in the most important points of a given experimental sequence in an experimental report independently and bringing data from measurements in the correct order.
Interpreting data	Drawing conclusions from everyday experiments.
	Drawing conclusions from experiments with respect to influencing factors, their variation and connected hypotheses.
	Drawing conclusions from experiments with respect to complex influencing factors, their variation and connected hypotheses.
Distinguishing science-related questions from other ones	Distinguishing science-related questions from other ones using given criteria
	Acknowledging criteria of science-related questions
	Distinguishing science-related questions from other ones with respect to suitable criteria
Acknowledging chances and risks of human behaviour	Acknowledging chances and risks of human behaviour in everyday life and drawing conclusions for responsible behaviour
	Acknowledging chances and risks of human behaviour in science-related contexts and drawing conclusions for responsible behaviour
	Acknowledging chances and risks of human behaviour in science-related contexts and drawing conclusions for responsible behaviour for oneself and the society

The second part covers inquiry competences. In the German competency model, this part is divided into three areas: scientific inquiry, scientific modelling and nature of science [32]. The Austrian competency models only cover the first area: scientific inquiry. In scientific inquiry, different actions can be distinguished. Pedaste et al. [33] identified a varying number of inquiry-related actions in their literature review. As example, two models are presented. Chinn and Malhotra [34] differentiate five different actions: generating a research question, planning an investigation, performing observations, explaining results and developing theories. Hofstein, Navon, Kipnis and Mamlok-Naaman [35] state similar phases: generating

research questions and hypotheses, planning an experiment, performing an experiment and analyzing data. The stated inquiry competencies follow the structure of the inquiry circle [33]. But so far, not all parts of the inquiry circle have been included in the IKM. For example asking questions, the starting point of the inquiry circle, has not been included in the IKM so far. In literature, empirically generated complexity levels for inquiry competences can be found for the German standards tests [20]. For each aspect – research questions, hypotheses, experimental design and analyzing data – in this standards test, five different levels were stated. Only on level one closed task designs are used (hypotheses: identifying one hypothesis which belongs to a described

experiment; experimental design: choosing a suitable experimental design for a simple investigation; analyzing data: single data can be identified). Tasks for all other levels used open-ended tasks, so they do not meet the requirements for IKM tasks. Hammann, Phan and Bayrhuber [36] used multiple choice items for evaluating the SDDS model (scientific discovery as dual search). The SDDS model states three different phases of inquiry: search in the hypotheses room, testing of hypotheses and analyzing evidence [37]. The authors presented multiple choice items for all three phases, but no graduation was used. Computerized testing of experimental competences was used by Schreiber, Theyßen and Schecker [38], who used simulations of experiments and recorded actions on the computer screen and answers on a protocol sheet. No instrument could be found in literature for inquiry competences and their graduation which only used closed task designs.

The third category of the competences – drawing conclusions – is represented by two competencies: identifying chances and risks of human behaviour and being able to distinguish science-related questions from questions in other contexts. The second one of these competences – distinguishing science-related questions from others – is part of the research about the nature of science. It deals among other topics with typical characteristics of nature science [39]. As it is normally assessed using assessment scales and not competency tasks, no complexity levels can be derived using literature about the nature of science. This competence is also listed in the fundamental abilities necessary to do science inquiry in grades 5 to 8 in the NRC standards [40]. In the German competency model, a related competence can be found in the “inquiry competences” section, which is divided into three aspects: doing inquiry, using models and reflections about the philosophy of nature science [41]. The last aspect is similar but not identical with the competence stated for the IKM.

The “identifying chances and risks of human behaviour” competence can partly be seen as component of the educational sustainability approach. Resources are limited and need to be dealt with responsibly. Different groups act with respect to variable goals and therefore stress the resources in different ways [42]. A common method used in this field of research is the Commons Dilemma approach [43]. It covers the steps problem diagnosis, policy decision making, practical interventions and effectiveness evaluation. For the IKM, especially the first two steps are crucial. On level one, the diagnosis of problems in everyday life is required, on level two the diagnosis of special science-related problems and on level three decisions with respect to different groups involved have to be made.

1.2. Assessment in Science-Related Classes

The term “science-related classes” is used here for classes in biology, chemistry and physics to distinguish them from science classes as in Austria, science classes can

only be chosen voluntarily. The Austrian guideline for assessment in class [44] contains two different kinds of classroom assessment. The first one is the assessment through written exams. Written exams contain tasks for different topics and competences, which are arranged by the teacher for each class individually. Mostly they contain open tasks about topics, students dealt with during the last weeks before the exam. In junior classes, they take ten minutes at most; in senior classes they are limited to 20 minutes. The second possibility is the use of alternative forms of assessment, like for example portfolios, examination reports, oral presentations or projects [45].

In addition, assessments for purposes, which are not directly part of science lessons (for example research or assessments for justifying political decisions), are used with students as well. Whereas classroom assessment is designed by teachers themselves, for the use in their classes, with not much thought to quality criteria like reliability, objectivity and validity, assessment tools for official purposes are developed with respect to quality criteria. They are mostly used by external persons and in many classes to produce the required sample size and cannot be used whenever a teacher needs them [46]. So on the one hand we have informal testing conditions in class and on the other hand we have formal testing conditions for external tests. Informal assessment refers to kinds of assessment conducted by teachers themselves for marking; formal assessment means nationwide official standards tests (which have been only available for mathematics, English and German so far). The results of official standards tests are reported to headmasters, school inspectors and the government, whereas results of informal assessment remain with the teachers [47].

With the introduction of science standards, assessment got more demanding for teachers. With respect to competency-orientated final exams, they have to assess not only the knowledge about the topics, which are given in curricula, but they also have to know about the competency levels their students have required so far, so they can make the right decisions for improving their students’ performance [12]. So teachers have to improve their formative classroom assessment competences [48]. Therefore informal methods of assessment may not be enough and new methods are needed that provide teachers with reliable information about their students’ competencies, but can be carried out by teachers easily, do not require special knowledge and can be used whenever it seems appropriate. These methods have to be situated somewhere between formal and informal testing procedures. Maybe it would be adequate to call them informal assessment plus [49].

The IKM is such an assessment tool between formal and informal testing (so far the only one that exists for science competencies in Austria) because it links typical characteristics of informal assessment, like the use in class for teaching relevant purposes, with characteristics of formal assessment, like the empirical evaluation of the tasks. Although big samples were used for evaluating the

underlying complexity levels and the diagnosis tasks, no sample norms are given to teachers as the IKM was not created for providing social comparison. Actually, it was developed for criteria-guided assessment. IKMs are not only created for science; they already exist for mathematics, English and German. So most teachers already know this tool and some have even used it for competency assessment before. The characteristics of IKMs are the computerized testing and the automatic evaluation of answers. Therefore the advantages are efficiency of time and objective testing. But due to these characteristics, some restrictions also have to be taken into account. For example, only closed task designs, like multiple choice tasks, cloze tests or numbering tasks could be used. This restriction in task design excludes some competences from being part in the IKM, like performing an experiment, and reduces the possibilities for some other competences, like naming or describing natural phenomena or depicting information.

As far as quality criteria are concerned, reduced standards have been evaluated for IKM as well. Validity makes sure that a test really measures what it pretends to measure. For tasks of the IKM this was ensured by expert rating by people who work at the universities of Salzburg and Vienna in the area of science education and are experienced teachers. Only tasks which were accepted in these groups were chosen for IKM. Objectivity was secured by using computerized testing and automatic evaluation of answers. Only reliability was not evaluated so far as no re-tests were conducted and no parallel-tests were used. But this is not astonishing at this stage of development. First the theoretical construct has to be confirmed, then items for the diagnosis instrument have to be created and subsequently, reliability analysis can be conducted. At the moment, the confirmation of the theoretical construct is in the center of attention. The other steps are going to be taken into account later. This is not unusual for this kind of study. Other studies at this stage of development do not deal with reliability analysis either [50,51]. Besides, studies which have already run through all steps of development, like the IQB-country comparison for competencies in mathematics and science in junior classes, do not report quality criteria, like reliability scores at all [20].

In this study, a clear distinction has to be made between the assessment to create an IKM for science and research questions. In our research, we wanted to address the following question: Can the theoretically proposed complexity levels be supported by empirical data? or in other words: Are items for level one easier to answer for the students than items for level two and are items for level two easier to answer than items for level three? If this question can be answered positively for a competence, the

underlying complexity structure (table 1) is confirmed for that competence. It is important to check the complexity levels first because evaluated complexity levels are the basis of an assessment instrument like the IKM. Producing a great amount of tasks cannot be started until these competency levels have been verified.

2. Materials and Methods

For answering our research question, a field study in school classes for 7th grade was carried out using closed tasks in a computerized testing environment. Testing took one school lesson (50 minutes) and was conducted by the teachers themselves. The test design is shown in figure 2. First of all, tasks for all competences and all competency levels were created. Therefore contexts of all three subjects were used. For each topic, one task containing three items, one for each level, was created. In a first step, an experienced teacher thought about an idea for the task. Then it was discussed in the subject-specific expert group containing at least five people: two people working in the area of science education at universities and three or four teachers with experience in the intended subject. For the next expert group meeting, the teacher created a task out of his idea, which was reworked in the expert group meeting. If the expert group of a subject was satisfied with a task it was presented to the expert groups of the other subjects for more detailed checking.

The next step was the testing of the tasks in class (pre-evaluation of tasks). In a first round, each task was tested in two classes. Then the results were presented in expert groups again. If necessary, the task was sent back to the revision team. If big revisions were necessary, the task was sent back to testing in class. Tasks which passed class testing were used in the main study. On the whole, 180 items were used in the main study, 87 items in study I and 93 in study II. 9 items were anchor items and 9 items were revised after main study I and used in main study II again. Booklets for study I contained 16 items, booklets for study II 21 because it got obvious that students could answer more than 16 items during one school lesson. Criteria for assignment of items to a booklet were the right competence and competency level mix (all competencies had to be in a booklet as well as a balanced level mix), estimated time for working on the task (from class evaluation) as well as the consideration of anchor items to link the results of both main studies. Exclusion criterion was the use of different levels of one task in the same booklet. Items were distributed to the booklets using the balanced incomplete block-design [52].

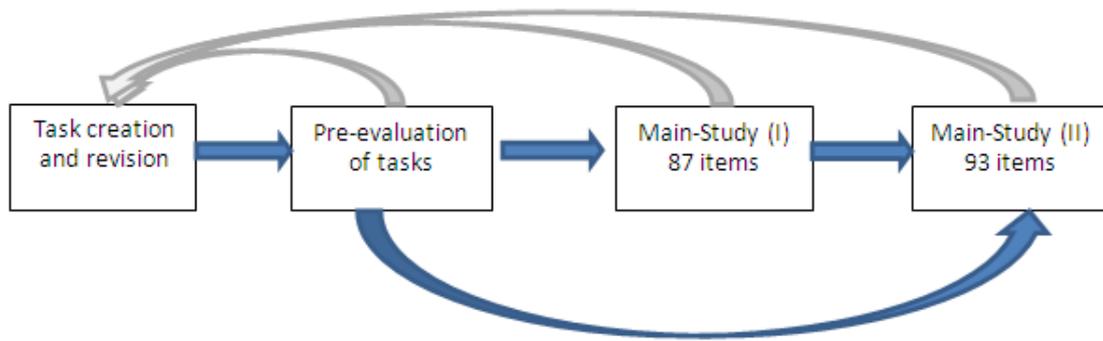


Figure 2. Test design

The sample consisted of students at the end of 7th grade, 2712 in study I and 1049 in study II. Schools were invited to take part in the study but could not get forced if they refused. Schools which decided to take part assigned classes for taking part in the assessment. After reporting the number of students, the teacher got codes for the testing platform – one code for each student. Testing took place under the supervision of the teacher during a normal school lesson. After the use, the code expired, so no changes could be made afterwards. Results were collected and evaluated automatically before being reported to researchers. Then the data underwent statistical evaluation using the Rasch-modelling software WINSTEPS 3.81.0 and IBM SPSS 21.0. Tasks were excluded if they were too difficult to answer (less than 5 % of the students could answer them) or if they were too easy (more than 95 % could answer them) or if tasks were more difficult to answer for students with other mother languages than German. For evaluation, the dichotomous logistic Rasch-Model was used. It was analyzed whether the tasks fit the calculated one-dimensional Rasch-Model. If that was not the case, they had to be excluded from further data evaluation as it had then to be assumed that the probability of answering the task right was determined not only by the intended competence. For the remaining tasks, difficulty scores were calculated. Difficulty scores show the difficulty of tasks on a logit scale: the smaller the score the easier a task is to answer for the sample. This procedure was necessary because not all tasks were answered by every single student. Instead they only got one booklet. WINSTEPS 3.81.0 is able to calculate difficulty scores even though not all students answered the same collection of tasks.

3. Results

For the "organising knowledge" category, three competences were evaluated: naming natural phenomena, describing natural phenomena and depicting information by using logical representations (table 2). For "naming natural phenomena", the competency levels ranged from using everyday language (level I) to naming phenomena with given scientific terms (level II) and using scientific terminology (level III). The first two columns of table 2

show the tasks, their levels and their difficulty scores for "naming natural phenomena". Scores range from minus infinite to plus infinite on a logit scale. The smaller the score the easier a task can be answered. Competency levels are confirmed if the score for the level I item is smaller than the score for the level II item and the score for the level II item is smaller than the score of the level III item. At the first glance, the theoretically proposed complexity levels get confirmed by task 2, 4 and 5. Tasks 1 and 3 do not show the required levels. When giving this a more detailed look, it gets clear that level II items of both tasks are easier to answer than level I items.

In columns three and four, results for "describing natural phenomena" are presented. The competency levels used here range from using everyday language (level I) to using science terminology (level II) and using underlying concepts for describing phenomena (level III). For task 1, no results are shown because this task does not fit the calculated one-dimensional Rasch-model.

Two of the tasks, task 3 and task 5, confirm the proposed competency levels whereas tasks 2 and 4 show a different graduation. For task 2, the level III item is easier to answer than the level II item and for task 4 the level II item is easier to answer than the level I item.

The last two columns of table 2 show the results for "depicting information by using logical representations". For this competence, diagrams and tables were used as logical representations. The competency levels can be extracted from table 1. On the whole, only task 5 confirms the proposed complexity structure. All other tasks show difficulties of various kinds. But no pattern is visible.

Table 3 shows the results of the competency level evaluation for inquiry competences. The upper part of the columns one and two contains the data for "formulating hypotheses". As can be seen, the first four tasks confirm the proposed competency levels, shown in table 1. Only task number 5 does not show the required graduation because the item for level II is easier to answer than the item for level I. But on the whole, the graduation is confirmed. The graduation of "planning an experiment" is also confirmed by all prepared tasks (upper part of columns two and three of table 3). The same is the case for "writing experimental records (lower part of columns one and two of table 3).

Table 2. Tasks for the “organizing knowledge” category with levels and corresponding difficulty scores (N = naming natural phenomena; Des = describing natural phenomena; DI = depicting information by using logical representations)

Task	Difficulty	Task	Difficulty	Task	Difficulty
N 1-I	0.04	Des 1-I		DI 1-I	- 1.58
N 1-II	- 0.19	Des 1-II		DI 1-II	1.83
N 1-III	2.34	Des 1-III		DI 1-III	0.81
N 2-I	- 1.64	Des 2-I	- 3.71	DI 2-I	- 2.30
N 2-II	- 0.68	Des 2-II	- 2.56	DI 2-II	- 3.81
N 2-III	4.02	Des 2-III	- 2.78	DI 2-III	0.83
N 3-I	- 1.41	Des 3-I	- 2.13	DI 3-I	3.13
N 3-II	- 3.80	Des 3-II	- 1.85	DI 3-II	1.50
N 3-III	- 1.30	Des 3-III	3.63	DI 3-III	5.46
N 4-I	- 1.69	Des 4-I	1.30	DI 4-I	- 2.71
N 4-II	- 0.15	Des 4-II	- 0.72	DI 4-II	- 0.51
N 4-III	5.12	Des 4-III	5.56	DI 4-III	- 0.57
N 5-I	- 2.56	Des 5-I	- 0.44	DI 5-I	- 1.52
N 5-II	- 0.65	Des 5-II	1.78	DI 5-II	0.09
N 5-III	2.56	Des 5-III	1.92	DI 5-III	1.74

Table 3. Tasks for the “gaining information through inquiry” category with levels and corresponding difficulty scores (H = formulating hypotheses; PE = planning an experiment; M = performing measurements; R = writing experimental records; ID = interpreting data)

Item	ISP	Item	ISP	Item	ISP
H 1-I	- 4.41	PE 1-I	- 0.34	M 1-I	- 1.23
H 1-II	- 2.40	PE 1-II	1.15	M 1-II	- 1.90
H 1-III	0.18	PE 1-III	2.64	M 1-III	1.64
H 2-I	0.31	PE 2-I	- 2.79	M 2-I	2.12
H 2-II	0.72	PE 2-II	0.60	M 2-II	0.49
H 2-III	1.90	PE 2-III	1.42	M 2-III	- 1.36
H 3-I	- 0.18	PE 3-I	- 1.19	M 3-I	- 1.23
H 3-II	0.17	PE 3-II	- 0.40	M 3-II	0.70
H 3-III	1.83	PE 3-III	2.46	M 3-III	- 0.93
H 4-I	- 4.41	PE 4-I	- 1.30	M 4-I	- 3.03
H 4-II	1.36	PE 4-II	- 0.96	M 4-II	- 2.30
H 4-III	1.58	PE 4-III	0.86	M 4-III	- 1.11
H 5-I	0.69	PE 5-I	- 1.42	M 5-I	- 0.71
H 5-II	0.55	PE 5-II	1.78	M 5-II	2.59
H 5-III	1.60	PE 5-III	1.92	M 5-III	6.26
R 1-I	- 3.02	ID 1-I	- 0.12		
R 1-II	- 1.34	ID 1-II	- 0.86		
R 1-III	2.73	ID 1-III	1.19		
R 2-I	- 1.42	ID 2-I	- 0.32		
R 2-II	2.56	ID 2-II	1.19		
R 2-III	4.67	ID 2-III	0.25		
R 3-I	- 2.40	ID 3-I	- 1.24		
R 3-II	- 2.25	ID 3-II	3.72		
R 3-III	- 0.72	ID 3-III	1.83		
R 4-I	- 1.42	ID 4-I	- 2.10		
R 4-II	0.42	ID 4-II	- 0.08		
R 4-III	2.18	ID 4-III	- 0.32		
		ID 5-I	- 3.60		
		ID 5-II	0.25		
		ID 5-III	- 0.10		

In columns five and six of table 3, empirical data for “performing measurements” are given. For this competence, the complexity levels are supported by all tasks. In fact tasks 4 and 5 show the required graduation whereas tasks 1 to 3 show different kinds of variations which are not consistent with the proposed competency levels. The last inquiry competence, interpreting data, is similarly problematic as empirical data do not support the proposed graduation. In most of the tasks, the problem is the difficulty of level III items. They are easier to answer than level II items. Only task 1 does not show this error; here, however, there are problems with level II being easier than level I. On the whole, the proposed graduation cannot be replicated.

For the “drawing conclusions” category, two different competences were examined: “distinguishing science-related questions from not science-related ones” and “acknowledging chances and risks of human behaviour to nature”. For both competences, empirical data show ambivalent results (table 4). For distinguishing science-related questions from other ones, tasks 3 and 4 show the proposed competency levels whereas tasks 1, 2 and 5 do not follow this graduation. The main problem of these three tasks is the difficulty of level I items as they are more difficult to answer for the sample than level II items. For acknowledging chances and risks of human behaviour, the tasks 3, 4 and 5 support competency graduation whereas tasks 1 and 2 do not. The problems however, are diverse. For task 1, the item for level I is more difficult to answer than the item for level II whereas for task 2, the item for level II is more difficult to answer than the item for level III.

Table 4. Tasks for the “drawing conclusions” category with levels and corresponding difficulty scores (SQ = distinguishing science-related questions from not science-related ones, CR = acknowledging chances and risks of human behaviour)

Item	ISP	Item	ISP
SQ 1-I	0.17	CR 1-I	0.96
SQ 1-II	- 1.23	CR 1-II	- 1.17
SQ 1-III	- 0.79	CR 1-III	1.79
SQ 2-I	0.59	CR 2-I	- 4.94
SQ 2-II	- 0.27	CR 2-II	2.25
SQ 2-III	2.86	CR 2-III	0.78
SQ 3-I	- 2.36	CR 3-I	- 2.32
SQ 3-II	- 1.10	CR 3-II	- 0.03
SQ 3-III	1.48	CR 3-III	1.79
SQ 4-I	- 1.09	CR 4-I	- 0.59
SQ 4-II	- 0.27	CR 4-II	2.57
SQ 4-III	4.43	CR 4-III	6.89
SQ 5-I	- 1.51	CR 5-I	- 7.25
SQ 5-II	- 2.22	CR 5-II	- 1.99
SQ 5-III	4.43	CR 5-III	6.08

Summing up the results draws a heterogeneous picture. For some competences, like formulating hypotheses,

planning an experiment and writing experimental records, the proposed competency graduation is supported by empirical data. For others, empirical data imply smaller changes in graduation or the restriction to special kinds of tasks. Here “naming natural phenomena”, “performing measurements” or “acknowledging chances and risks” can be named. The third category includes competences the proposed graduation of which requires major improvement or further testing, like depicting information by using logical representations, drawing conclusions from empirical data and distinguishing science-related questions from other ones.

4. Discussion and Conclusions

Standards for science education have been established in many English speaking countries for about thirty years now [39,53]. In German speaking countries, this tradition is much younger. Development only started about ten years ago [54, 5]. In Germany, Austria and Switzerland the development took place quite simultaneously. First, competency structure models were stated; then, competency development models were created. So far, these competency development models have been evaluated for many competences. In Germany, the first official standard test in mathematics and science took place [20] in 2012. In Austria, no official standard tests for science are planned at the moment. Nevertheless, the development of science standards in Austria is geared to developments in Germany and uses results from German research groups for developing science-related issues.

With regard to our research question: “Are items for level I easier to answer than items for level II and are items for level two easier to answer than items for level III?” we found different answers. For the “organizing knowledge” category, for two of the three competences - naming and describing natural phenomena - empirical data supported our theoretically proposed competency levels. These results are consistent with the empirically gained competency levels from the German nationwide competency test conducted in 2012. On the whole, five levels were stated for this test. On the easiest level, students were able to reproduce single biological facts in an everyday context. On level two, they were able to reproduce and explain simple biological relations, on level three, they could use concepts for simple explanations, on level four, they could use concepts for explaining complicated biological relations and on level five, students were able to explain relations that were unknown to them using concepts [20]. For the “naming natural phenomena” and “describing natural phenomena” competences, the complexity levels for the IKM followed the graduation from the German standards test: for level one, everyday vocabulary had to be used passively, which means different terms from everyday language were presented and students had to choose the correct one. For level two, science-specific terminology had

to be used and for level three, concepts had to be used for explanations (table 1). For “naming natural phenomena”, only the two physics tasks showed problems with complexity graduation. For both tasks, questions for level II were easier to answer than questions for level I. A closer look on the tasks which did not confirm the proposed competency levels showed that terms had to be used which could not be clearly allocated to everyday language or science terminology [21,22,23]. Terms like “energy” or “transformer” can be used in everyday contexts as well as in specific physical contexts. To straighten out graduation, the use of terms that are only used in everyday language is advisable. For open-ended items it is also recommendable to make a clear distinction between the passive use of everyday language, which means to choose from given language terms, the passive use of science-related terminology and the active use of science-related terminology (table 5).

Table 5. Recommendations for changed competency levels for “naming natural phenomena”

Competency	Levels
Naming natural phenomena	Naming natural phenomena by using given everyday language terms
	Naming natural phenomena by using given science-specific terms
	Naming natural phenomena by using science-specific terminology

For “describing natural phenomena”, two of the four tasks that corresponded to the Rasch-model confirm the proposed competency levels, whereas two do not follow the graduation. A closer look at these tasks reveals that both tasks use pictures that have to be described. Both tasks with correct graduation only use verbal information. Because verbal information and illustrations are different kinds of language [47], it is likely that they require different competency graduations. Taking these results into account, not the complexity levels have to be changed, but tasks have to be limited to using verbal information for making descriptions. Not many tasks are available from the German standards test. But the ones which were made public also used verbal information. In these tasks, pictures were only used for illustrating given verbal information but were not essential [20]. An instrument to assess the use of scientific language in biology classes also differentiates between verbal information, pictorial information and symbolic representations [47].

The third competency of knowledge organization deals with depicting information by using logical representations. Results from comparable studies only provide results from open-ended tasks [27,28,29,31]. However, Baker, Corbett and Koedinger used a multiple choice item for selecting the correct type of graph [29]. According to the results found in literature, we stated complexity levels for the use in closed tasks: ranging from depicting single data points within a given frame to acknowledging the main characteristics of a representation to executing all three steps mentioned in

literature: choosing a correct type of representation, creating the frame and entering data points into the frame. However, only two tasks show the required graduation. The main problem here is the use of different kinds of logical representations. Whereas in literature only graphs are used [36,27,28], we also used tables in our tasks. So the first step has to be the limitation to one kind of logical representation. In our case we choose graphs as they are better described in literature. For the work with graphs in class, two different types of competencies can be distinguished: competences for depicting information and competences for extracting information. As extracting information is much easier to realize using closed task designs, we first want to concentrate on this part of diagram competence. According to literature, we propose the following graduation for “extracting information from diagrams” [27,37,38] (table 6).

Table 6. Recommendations for competency levels for “extracting information from diagrams”

Competency	Levels
Extracting information from diagrams	Identifying the content shown in a diagram
	Extracting information of a single data point
	Comparing information of data points

For the “gaining information through inquiry” category, five different competences were stated, following the inquiry circle: formulating hypotheses, planning an experiment, performing measurements, writing experimental records and interpreting experimental data [26,59]. In literature, studies could be found either with respect to graduation of inquiry competences [29] or using closed task designs [36]. As no graduation could be found in literature for closed task designs and the ones for open-ended tasks were not appropriate, competency graduation had to be stated without help from literature. For “formulating hypotheses”, empirical data confirm the proposed complexity levels, so no changes need to be made. The same is the case for “planning an experiment”. So there is no need for changing competency levels. For “writing experimental records”, no changes need to be made either. A different situation can be found for “performing measurements”, where only two tasks show the proposed graduation: reading data from a measurement instrument, performing a measurement and considering limits for measurement. A more detailed look reveals the possible reason: tasks 1, 2 and 3 are biology tasks concerning counting as kind of measurement. For biology, doing measurements is not as essential as for physics and chemistry. Besides, more common types of measurements, like temperature or weight measurements are used. Task 4, a physics task, shows the required competency levels as well as task 5 – a chemistry task. But task five is quite difficult to answer for students as when they were doing the study, students had not had any lessons in chemistry and therefore had to rely on knowledge from physics. As measurements are not as crucial for biology, a possible solution is the restriction of this competence to physics and chemistry. But

more tasks need to be created for these two subjects to confirm the proposed graduation. Major problems can also be found for “interpreting data”, where not a single task shows the required competency levels. The main problem is the difficulty of level III items as in four tasks they are easier to answer than level II items. A closer look at the tasks reveals that these level III items contain diagrams which depict two influencing factors. Two problems arise from this approach. First, diagram competences play a role in these tasks. This is problematic as for measurement, it is not advisable to mix up competences. Second, the diagrams simplify the depiction of the complex influencing factors, so they do not seem as complicated any more. To solve this problem, the influencing factors need to be depicted without diagrams, in a way that makes complexity more obvious, maybe by describing the interaction of the factors verbally.

For the “drawing conclusions” category, two different competences are part of the IKM: distinguishing science-related questions from other ones and acknowledging chances and risks of human behaviour to nature. Both competences are quite specific for the Austrian competency model. The identification of science-related questions and the justification why some questions are science-related and others are not, are subject of the research about the nature of science. Assessment scales and not competency tasks are the most common type of assessment methods in this field [58]. A related competence can be found in the German competency model in the inquiry competences section where one part deals with the reflection of the philosophy of science inquiry [41]. Unfortunately, no clear description of this competence and no task examples are included, so we were on our own for stating competency levels for the IKM. For distinguishing science-related questions from other questions, level I used given criteria for the distinction; on level II, some criteria for distinction had to be named and on level III, the distinction had to be made by self-selected criteria. For the tasks with biological and physical context, the proposed competency levels were confirmed by empirical data. For the tasks using chemistry contexts, level I items were more difficult to answer than level II items. A possible explanation is that the students had not had lessons in chemistry when they were answering the study. This is due to the fact that in Austria, there are no chemistry lessons until 8th grade. Empirical data do not show that chemistry tasks for this competence are generally more difficult to answer than items for other contexts but maybe task writers gave more cues for answering the items right and therefore complexity levels got mixed up. It would be best to check the tasks in that respect and to test chemistry items again, but at the end of 8th grade, when students had had chemistry lessons. For “acknowledging chances and risks of human behaviour”, which is partly a component of the environmental sustainability approach [42,43], one way of investigating this area are commons dilemmas. Commons dilemmas get more complicated the more actors are involved. For the IKM, the involvement of global actors is part of level III. On level II, the distinguishing criterion is the familiarity

of the problem to students (everyday versus science-related problems). In the German competency model, it would be part of the “valuation” section, but this section has not been part of the national standards test so far and therefore no empirical data about complexity levels are available. As far as the evaluation of the graduation for the IKM is concerned, three tasks show the required competency levels, but two tasks do not support the proposed graduation. For task 1, the item for level I is more difficult to answer than the item for level II. For task 2, the item for level II is more difficult to answer than the item for level III. For both tasks, the problems are due to special characteristics of the problematic items. Therefore both items should be overworked and tested again.

On the whole, the study shows that for many competences like formulating hypotheses, planning an experiment and writing experimental records, the proposed competency graduation does not need further changes. For others, empirical data imply smaller changes in graduation or the restriction to special kinds of tasks. Only for “depicting information by using logical representations”, complexity levels have to be changed completely.

REFERENCES

- [1] Organisation for Economic Cooperation and Development (OECD). Programme for international student assessment (PISA), Results from PISA 2012, Country Note: Austria, Online available from <http://www.oecd.org/austria/PISA-2012-results-austria.pdf>
- [2] Organisation for Economic Cooperation and Development (OECD). PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science, 2010, , Online available from <http://dx.doi.org/10.1787/9789264091450-en>
- [3] U. Schwantner, U., C. Schreiner, (Hrsg.). PISA 2009: Internationaler Vergleich von Schülerleistungen: Erste Ergebnisse Lesen, Mathematik, Naturwissenschaft, Graz, Leykam, 2010.
- [4] H. Stadler, A. Lembens, H. Weiglhofer. PISA Naturwissenschaft: Die österreichischen Ergebnisse. In C. Schreiner, U. Schwantner, (Hrsg.), PISA 2006: Österreichischer Expertenbericht zum Naturwissenschaftsschwerpunkt, Graz, Leykam, 2009.
- [5] Federal Institute of Research, Innovation and Development of the Austrian school system (BIFIE). Bildungsstandards in Österreich: Überprüfung und Rückmeldung. 4. Aktualisierte Auflage, 2012, Online available from <https://www.bifie.at/no/de/560>
- [6] Federal Institute of Research, Innovation and Development of the Austrian school system (BIFIE), NAWI-Bildungsstandards: Aktueller Stand und Ausblick, 2012, Online available from <https://www.bifie.at/node/2076>
- [7] Bundesgesetzblatt für die Republik Österreich. 117. Bundesgesetz, Teil I: Änderung des Schulunterrichtsgesetzes, ausgegeben am 8. August 2008.

- [8] Bundesgesetzblatt für die Republik Österreich. Teil II: Verordnung über Bildungsstandards im Schulwesen, ausgegeben am 2. Jänner 2009.
- [9] Bundesgesetzblatt für die Republik Österreich. 174. Verordnung, Teil II: Verordnung der Bundesministerin für Unterricht, Kunst und Kultur über die Reifeprüfung in den allgemeinbildenden höheren Schulen (Prüfungsordnung AHS), ausgegeben am 30. Mai 2012.
- [10] F. E. Weinert, F. E. Concepts of Competence. Definition and Selection of Competencies. Paris, OECD, 1999.
- [11] I. Venus-Wagner, H. Weiglhofer, J. Zumbach. Kompetenzorientiertes Unterrichten in den Naturwissenschaften. In: M. Paechter (Hrsg.), Handbuch Kompetenzorientierter Unterricht. Weinheim, Beltz, 188 – 202, 2012.
- [12] I. Schiffll, H. Weiglhofer. Kompetenzdiagnostik im Biologieunterricht am Beispiel des Instruments zur informellen Kompetenzmessung. In G. Maresch & J. Zumbach (Hrsg.), Didaktik der Naturwissenschaften: Neu Horizonte in Biologie, Geometrie und Informatik, Wien, Facultas, 27-41, 2016.
- [13] I. Schiffll, H. Weiglhofer, D. Remplbauer. Development of inquiry competencies during junior classes: A replication study using longitudinal design. *Science Education Research Letters*, Volume 2015, 16 -20, 2015, Online available from <http://edoc.hu-berlin.de/serl/2015/3/PDF/2015-3.pdf>
- [14] K. Neumann, A. Kauertz, A. Lau, H. Notarp, H. E. Fischer. Die Modellierung physikalischer Kompetenz und ihre Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 105-121, 2007.
- [15] I. Venus-Wagner. Kompetenzorientierte Aufgaben, *Der mathematische und naturwissenschaftliche Unterricht*, 64, 428 – 432, 2011.
- [16] Bundesinstitut für Forschung, Innovation und Entwicklung des österreichischen Schulsystems (BIFIE). Kompetenzmodell Naturwissenschaften 8. Schulstufe, 2011, Online available from https://www.bifie.at/system/files/dl/bist_nawi_kompetenzmodell-8_2011-10-21.pdf
- [17] M. Walpuski, N. Kampa, A. Kauertz, N. Wellnitz, N. Evaluation der Bildungsstandards in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht*, 61, 323-326, 2008.
- [18] M. Walpuski. et al. ESNaS-Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In: A. Gehrmann, U. Hericks M. Lüders (Hrsg.): Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht Bad Heilbronn, Klinkhardt, 171-184, 2010.
- [19] J. Grünkorn, A. Upmeyer zu Belzen, D. Krüger. Assessing students' understanding of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 36, 1-34, 2014.
- [20] J. Mayer, J. Wellnitz, N. Klebba, N. Kampa., Kompetenzstufenmodelle für das Fach Biologie. In H. Anand Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), IQB-Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I, Münster, Waxmann, 74-82, 2013.
- [21] H.R. Fluck. Fachsprachen: Zur Funktion, Verwendung und Beschreibung eines wichtigen Kommunikationsmittels unserer Gesellschaft. In K. Eichhoff-Cyrus, R. Hoberg (Hrsg.), *Die deutsche Sprache zur Jahrtausendwende: Sprachkultur oder Sprachverfall?*. Mannheim, Dudenverlag, 2000.
- [22] K. Rincke. Von der Alltagssprache zur Fachsprache - Bruch oder schrittweiser Übergang?. In G. Fenkhart, E. Zeitlinger, A. Lembens (Hrsg.), *Sprache, Mathematik und Naturwissenschaften*. Innsbruck, Studien Verlag, 2010.
- [23] T. Roelcke. *Fachsprachen*. Berlin, Erich Schmidt Verlag, 2010.
- [24] J. L. Lemke. *Talking Science: Language learning and value*. Norwood, NJ; Ablex Publishing Corporation, 1990.
- [25] J. Wellington, J. Osborne. *Language and literacy in science education*. Buckingham, Philadelphia, Open University Press, 2001.
- [26] B. Laufer, T. S. Paribakht. The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48, 365-391, 1998.
- [27] S. Lachmayer, C. Nerdel, H. Prechtel, H. Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 161-180, 2007.
- [28] L. von Kotzebue, C. Nerdel. Modellierung und Analyse von Professionswissen zur Diagrammkompetenz bei angehenden Biologielehrkräften. *Zeitschrift für Erziehungswissenschaft*, 18, 687-712, 2015.
- [29] R. S. Baker, A. T. Corbett, K. R. Koedinger. Toward a model of learning data representations. In J. D. Moore, K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Mahwah, NJ, Erlbaum, 45-50, 2001.
- [30] M.J. Waverling. Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching*, 26, 373-379, 1989.
- [31] D. Kerslake. Graphs. In K. Hart, D. Kerslake, M. L. Brown et al. (Eds.). *Childrens' understanding of Mathematics*, 11-16.
- [32] N. Wellnitz, H.E. Fischer, A. Kauertz, J. Mayer, I. Neumann, H. Anand Pant, E. Sumfleth, M. Walpuski. Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 261 – 291. 2012
- [33] M. Pedaste et al. Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47 – 61, 2015.
- [34] C. A. Chinn, B. A. Malhotra. Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86, 175-218, 2002.
- [35] A. Hofstein, O. Navon, M. Kipnis, R. Mamlok-Naaman. Developing students' ability to ask more and better questions resulting from inquiry type laboratories. *Journal of Research in Science Teaching*, 42, 791-806, 2005.
- [36] M. Hammann, T. T. Pahn, H. Bayrhuber. Experimentieren als

Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen?. Zeitschrift für Erziehungswissenschaften, 10, Sonderheft 8/2007, 33-49, 2007.

- [37] D. Klahr. Exploring science: The cognition and development of discovery processes, Cambridge.
- [38] N. Schreiber, H. Theyßen, H. Schecker. Experimentelle Kompetenzen messen?!. Physik und Didaktik in Schule und Hochschule, 83, 92-101, 2009.
- [39] National Research Council. National science education standards. Washington DC, National Academy Press, 1996.
- [40] National Research Council. Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC, National Academy Press, 2010.
- [41] E. Sumfleth et al. Das Kompetenzstrukturmodell in den naturwissenschaftlichen Fächern. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I, Münster, Waxmann, 38-42, 2013.
- [42] C. Vlek, L. Steg. Human behaviour and environmental sustainability: Problems, driving forces and research topics. Journal of Social Issues, 63, 1-19, 2007.
- [43] C. Vlek, L. Steg. The commons dilemma as a practical model for research and policy making about environmental risks. In G. Bartels, W. Nelissen (Eds.), Marketing for sustainability. Towards transactional policy-making. Oxford: IOS Press, 205-303, 2002.
- [44] Bundesgesetzblatt für die Republik Österreich. BGBl 371/1974. Verordnung über die Leistungsbeurteilung in Pflichtschulen, sowie mittleren und höheren Schulen (Leistungsbeurteilungsverordnung). Ausgegeben am 24. Juni 1974.
- [45] M. Birenbaum, F. Dochy (Eds). Alternatives in assessment of achievements, learning processes and prior knowledge. Springer Science and Business Media, Vol. 42, 2012.
- [46] F. E. Weinert (Hrsg.). Leistungsmessung in Schulen. Weinheim, Beltz, 2014.
- [47] S. Nitz, C. Nerdel, H. Prechtl. Entwicklung eines Erhebungsinstrumentes zur Erfassung und Verwendung von Fachsprache im Biologieunterricht. Zeitschrift für Didaktik der Naturwissenschaften, 18, 117 – 139, 2012.
- [48] J. H. McMillan. Formative Classroom Assessment: The key to improving student achievement. In J. H. Mc Millan (Ed.). Formative Classroom Assessment: Theory and Practice. New York, Teachers College Press, 1-7, 2007.
- [49] F.-W. Schrader, A. Helmke. Alltägliche Leistungsbeurteilung durch Lehrer. In F. Weinert (Hrsg.), Leistungsmessung in Schulen, 2014.
- [50] J. Grünkron, Upmeyer zu Belzen, A., D. Krüger. Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden, G. S. Carvalho (Eds.), Authenticity in Biology Education: Benefits and Challenges. A selection of papers presented at the 8th Conference of European Researchers in Didactics of Biology (ERIDOB), 2011.
- [51] S. Lachmayer. Entwicklung und Überprüfung eines Strukturmodells der Diagrammkompetenz für den Biologie-Unterricht. Dissertation an der Christian-Albrechts-Universität zu Kiel, 2008.
- [52] J. M. Gonzalez, J. L. Eltinge. Multiple Matrix Sampling: A review. Proceedings of the Section Survey Research Methods. American Statistical Association, 3069-3075, 2007.
- [53] Department of Education and Skills & Qualification and Curriculum Authority. Science – the national curriculum for England. HMSO, London, 2004.
- [54] KMK (Kultusministerkonferenz). Bildungsstandards im Fach Biologie für den mittleren Schulabschluss. 2005.
- [55] B. Winn. Charts, graphs and diagrams in education materials. In H. A. Houghton, D. M. Willows (Eds.). The psychology of Illustration, New York, Springer, 152-198, 1987.
- [56] H. Wainer. Understanding graphs and tables. Educational Researcher, 21, 14-23, 1992.
- [57] F. R. Curcio. Comprehension of the mathematical relationships expressed in graphs. Journal for Research in Mathematics Education, 18, 382-393, 1987.
- [58] N. G. Lederman. Students' and teachers' conceptions of the nature of science: A review of research. Journal for Research in Science Teaching, 29, 331-359, 1992.
- [59] N. Wellnitz, H.E. Fischer, A. Kauertz, J. Mayer, I. Neumann, H. Anand Pant, E. Sumfleth, M. Walpuski. Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. Zeitschrift für Didaktik der Naturwissenschaften, 18, 261 – 291. 2012