Science Education
Review Letters

Development of inquiry competencies during junior classes: A replication study using longitudinal design

Iris Schiffl[a*], Hubert Weiglhofer[a] and Daniela Remplbauer[a]
[a]School of Education, Universität Salzburg, Salzburg, Österreich

[*] Corresponding author: iris.schiffl@sbg.ac.at

Received 25th February 2015, Accepted 5th November 2015

Abstract
Scientific inquiry plays a major role in science education. So far competency structures related to knowledge gain through inquiry have been identified and analyzed. Different levels of such structures have also been articulated. In a nationwide Austrian study conducted by BIFIE (Bundesinstitut für Bildungsforschung, Evaluation und Entwicklung des österreichischen Schulwesens) competency levels were tested in order to construct a diagnosis instrument for natural sciences to diagnose junior class students. In the presented study we want to replicate the findings of the BIFIE study but not in a cross-sectional but in a longitudinal study from 5th to 7th grade. This study design was chosen because we wanted to answer the question how inquiry competencies develop during junior classes and the BIFIE study which was only conducted once in 7th grade could only give cues about the competency level of 7th graders but not about competency development.

Keywords
Science standards, inquiry competences, competency diagnosis

Introduction and Theoretical Background

Experiments are an important way of gaining knowledge in natural science. This is reflected in commonly introduced competency structure models (KMK, 2005; Walpuski et al., 2008; Venus-Wagner, Weiglhofer & Zumbach 2012). These models divide scientific inquiry into four domains: research questions, hypotheses, experimental design and data analysis (Mayer, Grube & Möller, 2008). After evaluating competency structure models empirically the next step is the construction of levels of competency for each domain (Mayer & Wellnitz, 2014). In accordance with the ESNaS competency structure model, competency levels for scientific inquiry were described using two different dimensions (complexity and cognitive processes) (Wellnitz et al., 2012). In Austria, another way was chosen. Instead of using the same complexity features for all competencies of acting, three different levels of complexity were expressed for each operator (see table 1).
Based on the competency levels in table 1, tasks were developed by experts and were tested by BIFIE in nationwide field tests. The best performing tasks were chosen for inclusion in a competency diagnosis instrument (in preparation by BIFIE: www.bifie.at)

Research Design and Method of BIFIE´s sectional design study

The BIFIE sample consisted of 3378 students of 7th grade from all over Austria. About 60% of the students attended grammar school. Testing took place in two computerized field tests. The BIFIE study not only covered inquiry competences but also the other two domains of acting (knowledge and evaluation) of the Austrian competency structure model (Venus-Wagner, Weiglhofer & Zumbach, 2012). For each domain of inquiry five tasks were developed. For each task three different multiple choice items were constructed - one for each level of the model (table 1). So on the whole 25 tasks with 75 items were used for diagnosing inquiry competences. Each student had to answer a booklet containing 12 of those 75 items.

For evaluation IRT (Item-Response-Theory) was used, because using IRT it is possible to guess students answers on rasch-scaled tasks even if

Tab 1: Competency levels for domains of scientific inquiry (Operators)

Operator	Levels
Disposing Hypotheses	Disposing hypotheses from a given issue
	Disposing hypotheses from a given issue with respect to influencing factors
	Disposing hypotheses from a given experiment with respect to influencing factors and their variaton
Planning an Experiment	Reproducing an experiment using given materials
	Reproducing an experiment without given materials
	Planning an experiment independently
Doing Measurement	Reading data from a given measuring instrument
	Choosing a suitable measuring instrument and doing measurement
	Choosing a suitable measuring instrument and doing measurement, consider limits of measurement and change between measurement units
Writing Records	Filling in the most important points of a given experimental sequence in a given experimental report (right order)
	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 measurement in a correct order.
Drawing Conclusions from Experiments	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

they had not actually answered them. For evaluation WINSTEPS 3.81.0 was used. From the computerized collected data (answers on items) a one dimensional rasch-model was calculated. Then it was examined which items did not fit the calculated model (using infit and outfit scores). Tasks were excluded from the BIFIE study if they did not fit the calculated one-dimensional RASCH-Model (for example tasks on which poor performers performed better than well performers) or results indicated other problems (for example problems in understanding). Afterwards using the remaining items a Wright-Map was constructed. A Wright-Map map links the answering probability of items with students´ performance. Items which could be answered by many students – easy items -are shown on the bottom of the logit scale. That is also the case for students who could not answer many tasks correctly. The better a student´s performance the higher he is situated on the scale. The theoretically proposed levels (table 1) are confirmed empirically if level I of a task was answered correctly more often (was shown lower on the logit scale of the Wright Map), than level II and level III.

Objective, Research Questions and Hypotheses

The presented study aims to replicate the purpose of the BIFIE study for examining competency levels as basis for a competency diagnosis instrument. For our study we chose only those items from the BIFIE study which empirically matched the theoretically proposed levels. Because we wanted to learn about the development of inquiry competencies during junior classes a longitudinal design was used in which students in different grades were examined.

We aimed to address the following question:

- Do inquiry competencies increase with age during junior classes?

Since we also tested 7th graders, we expected the frequency of right answers to be on a comparable level with the BIFIE study. For the lower class levels (5th grade and 6th grade) we expected them to be somewhat lower. On the whole a competency improvement from 5th to 7th grade was anticipated because in all three grades students had biology lessons and according to the Austrian competency model inquiry competencies should play an important role in science teaching. So more biology education should lead to improved inquiry competencies

Research Design and Method

Research Design and Method of longitudinal study

Our sample consisted of 384 students attending grammar school (53,6 %) or secondary modern school (46,4 %). Six different Upper Austrian schools took part in the study (three grammar schools and three secondary modern schools). The sample was divided approximately equally between 5th grade (32,8 %) , 6th grade (34,6 %) and 7th grade (32,6 %) students. About half of the students were girls (46,9 %), and half were boys (53,1 %).

The items for this study were taken from the BIFIE study. As stated before we chose only items which matched the theoretically proposed competency level empirically. We chose 20 multiple choice items. All items at level 1 and 2 were one-out-of-four multiple choice items, items at level 3 were often two-out-of-five multiple-choice items. Since all 20 multiple choice test items could not be answered by one student because time was limited to one lesson two different test booklets were created. As 5th graders are slower readers, as we expected their competencies to be on a more basic level and as we did not want to set students 'motivation at risk by asking too difficult questions, the booklets for 5th grade only contained the tasks on levels 1 and 2 - 8 items per booklet. Booklets for 6th and 7th grade contained all tasks on the three levels - 10 items per booklet. Testing took place during normal lessons, but was conducted by a qualified external person. The pupils had 50 minutes (one lesson) to answer the multiple choice items using paper and pencil. Afterwards the booklets were collected and analyzed using IBM SPSS 21.0

Results

In order to answer our research question - if inquiry competencies increased with age during junior classes - we first analyzed how many students of each class gave the correct answer to the presented tasks (table 1, columns "5th grade", "6th grade" and "7th grade", answers in %). It soon became obvious that in most tasks 7th graders did not perform any better than students of 5th or 6th grade. Actually, in many tasks

Table 2: Performance of students of different class levels (H = disposing hypotheses, P = planning an experiment, M = doing measurements, R = writing references, C = drawing conclusions), data in %

Operator /level	*Task*	*Reference (BIFIE study, 7th graders)*	*5th grade*	*6th grade*	*7th grade*	*Differences between classes*
H/1	Herzschlag	93,9	85,7	82,0	75,2	χ^2= 4,62; p=.10
H/1	Kerze	86,2	86,7	78,2	80,8	χ^2= 1,91; p=.39
H/2	Herzschlag	61,3	54,8	52,6	44,0	χ^2= 3,26; p=.20
H/3	Schimmelpilz	55,4		59,7	81,7	χ^2= 7,23 p≤.01
P/1	Axels Asseln	53,3	65,2	73,1	61,3	χ^2= 2,13; p=.35
P/2	Babyflasche	57,2	78,8	71,6	69,4	χ^2= 1,16; p=.45
P/3	Axels Asseln	17,7		19,7	15,9	χ^2= 0,32; p=.57
P/3	Tomaten	20,5		11,9	9,7	χ^2= 0,17; p=.68
M/1	Meerschweinchen	86,1	77,3	88,1	75,8	χ^2= 3,73; p=.16
M/2	Gänseblümchen	49,5	48,3	42,4	42,9	χ^2= 0,54; p=.76
M/2	Bodenproben	56,4	42,4	46,3	58,1	χ^2= 3,36; p=.19
M/3	Bakterienkultur	41,7		46,3	35,5	χ^2= 2,68; p=.26
R/1	Lieblingsfutter	86,0	86,4	82,1	87,1	χ^2= 0,75; p=.69
R/2	Fette	73,3	46,7	75,8	68,5	χ^2= 0,92; p=.93
R/2	Alufolie	56,9	71,7	77,3	65,1	χ^2= 2,34; p=.31
R/3	Lieblingsfutter	20,8		15,2	11,1	χ^2= 0,46; p=.50
C/1	Rostiger Nagel	85,0	81,8	83,6	79,0	χ^2= 0,45; p=.80
C/2	Samenkeimung	58,7	65,2	59,7	54,8	χ^2= 1,41; p=.49
C/2	Pflanzenwachstum	77,9	63,3	71,2	60,3	χ^2= 1,79; p=.41
C/3	Rostiger Nagel	66,2		69,7	71,4	χ^2= 0,05; p=.83

they performed equally or even worse. To test if there were any statistically significant differences between class levels we used the Kruskal-Wallis H-Test. The results are shown in table 2, column "Differences between classes".

As far as the correctness of answers is concerned the results of the Kruskal-Wallis Test show that students of different class levels differ in only one task (H3 - Schimmelpilz). In all other 19 tasks there are no significant differences between class levels. Due to these results, H1 – which states the differences between grades – has to be rejected.

Discussion and Outlook

Since the sample of the BIFIE study only consisted of 7th graders, no predication could be made about the development of competencies during junior classes. Therefore we chose students from different junior classes (5th to 7th grade) to examine the development of inquiry competencies during these years of school. We expected the inquiry competencies to increase with age because in all examined grades students did have biology lessons and as stated in the Austrian competency model for junior classes inquiry competencies should play a major role in teaching natural science (Venus-Wagner, Weiglhofer, Zumbach, 2012). But by contrast to our prediction, students´ competencies did not improve from 5th to 7th grade. A possible explanation could be that during the first years of junior classes there actually is little encouragement to do experiments in biology although the competency model states the importance of doing inquiry. The topics of the biology curriculum often do not invite experimental testing and many teachers focus on 'giving information' in these early years of science education often using question guided teaching for developing knowledge (Weiglhofer, Stadler & Lembens, 2009). Besides if a teacher uses experiments in these grades he often conducts them by himself and the students only have to watch (Grafendorfer & Neureiter, 2009). Furthermore physics only starts in 6th grade with one hour per week and chemistry is not taught at all in these age groups (BMBF, 2015). In both subjects experiments traditionally play an important role. If experimental work is not frequent, it is not surprising that students´ performance does not improve. At first glance the results give the impression that students´ experimental competencies are not bad, but when we consider a default 25% chance of scoring the correct answer, the data do not appear to reveal a group of students with strong competency.

On the whole the longitudinal study gives an impression of a possible development of science inquiry competencies in junior classes. The lack of improvement in inquiry competencies during junior classes leads to the conclusion that aspects of inquiry need to be practiced in biology lessons more often, so that students can improve and develop higher competency levels. However, this study also has its limitations: the sample is quite small, it is selected and not representative and only multiple choice items are used. Moreover 5th graders did not get level 3 items. Concerning our results, items of all levels should be used for all class levels because 5th graders did not score worse and so no motivational challenge caused by too difficult items should be expected. Furthermore using all items for all class levels could lead to further results about competency development.

The results of the study indicate that more training in science inquiry is necessary in junior classes. But to get a clearer view of the development of different aspects of inquiry competencies it is necessary to use a wider range of tasks (hands on, open, multiple choice and forced choice) because with multiple choice items it is not possible to examine all aspects of inquiry (for example perform experiments or really write a record). Therefore a follow up study on a larger sample should be taken into account to justify these first results of competency development by using a wider range of tasks.

References

Bundesministerium für Bildung und Frauen (BMBF, 2015). Stundentafeln und Lehrpläne der AHS Unterstufe. https://www.bmbf.gv.at/schulen/unterricht/lp/lp_ahs_unterstufe.html [22.Juni 2015]

Grafendorfer, A. & Neureiter, H. (2009). Unterricht in Naturwissenschaft. In C. Schreiner & U. Schwantner (Hrsg.), *PISA 2006: Österreichischer Expertenbericht zum Naturwissenschaftsschwerpunkt*. Graz: Leykam.

KMK (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München, Neuwied: Luchterhand.

Mayer, J., Grube, C. & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In: U. Harms & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik. Band 3*. Innsbruck: Studien Verlag.

Mayer, J. & Wellnitz, N. (2014). Entwicklung von Kompetenzstrukturmodellen. In: D. Krüger, i. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschafts-didaktischen Forschung*. Berlin: Springer Verlag.

Venus-Wagner, I. (2011). Kompetenzorientierte Aufgaben, *MNU, 64*, 428 – 432.

Venus-Wagner, I., Weiglhofer, H. & Zumbach, J. (2012). Kompetenzorientiertes Unterrichten in den Naturwissenschaften. In: M. Paechter (Hrsg.), *Handbuch Kompetenzorientierter Unterricht*. Weinheim: Beltz.

Weiglhofer, H., Stadler, H. & Lembens, A. (2009). Unterricht in Naturwissenschaft: Österreichische Ergebnisse aus fachdidaktischer Sicht. . In C. Schreiner & U. Schwantner (Hrsg.), *PISA 2006: Österreichischer Expertenbericht zum Naturwissenschaftsschwerpunkt*. Graz: Leykam.

Walpuski, M., Kampa, N., Kauertz, A. & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *MNU, 61*(6), 323-326.

Wellnitz, N. et al. (2012). Evaluations der Bildungsstandards: eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *ZfDN, 18*, 261-291. [http://archiv.ipn.uni-kiel.de/zfdn/pdf/18_Rabe.pdf]