LOOK ACADEMIC PUBLISHERS
OPEN ACCESS

INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
e-ISSN: 1306-3065
2018, Vol. 13, No. 8, 647-660

A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts

Margarethe Buse [1*], Karsten Damerau [1], Angelika Preisfeld [1]

[1] *University of Wuppertal, Department of Zoology and Biology Education, GERMANY*

* CORRESPONDENCE: mabueu@gmx.de

ABSTRACT

To date research on content and language integrated learning (CLIL) in scientific out-of-school programmes has been seldom been carried out (Rodenhauser & Preisfeld, 2015). The programme in question aims to provide learners with both, authentic scientific education and subject-related learning in the school language German and the first foreign language, English, to prepare for vocational and university training. With a longitudinal, quantitative, quasi-experimental design, the present study has explored to what extent teaching Biology in a CLIL intervention with English as the foreign language affects the acquisition of scientific knowledge and two experimentation-related ability self-concepts. For the investigation of this question, a pre-post-follow-up-design has been conducted with 170 senior students being taught CLIL in English and a control group of 48 students being instructed in German. From the pre test, self-assessments of affective data were employed to subgroup as primarily interested in sciences (*scientists*), as primarily interested in foreign languages (*language lovers*) and as interested in both (*all-rounders*). Regarding the cognitive knowledge, both, significant short-term and long-term increases were measured. Furthermore, cognitive acquisition proved to be independent of the treatment as well as of preference subgroups. For the ability self-concepts, however, group-depending differences were identified. While in the ability self-concept on experimenting a homogeneous, both, treatment- and preference group-independent short-term rise was registered, for the ability self-concept on interpreting data a treatment-dependent, group-independent, short-term decrease was observed. As to the cognitive gain and the development of self-concepts observed in this study, practical experimentation combined to a CLIL format can be assumed to be beneficial to heterogeneous learner groups.

Keywords: CLIL, experimenting, academic achievement, academic self-concepts, interpreting data, competence in experimenting, scientific literacy, out-of-school scientific programme, bilingual education

INTRODUCTION

As issues of general interest in society are very often discussed in English, the European Commission´s policy aims at all learners having a command both, of their native language and two foreign languages at a high functional level (European Commission, 2004). Moreover, English is regarded to be the language of science in Biology and other sciences being both, the predominant working and publishing language (Gnutzmann, Jakisch, & Rabe, 2015). Education based on content and language integrated learning (CLIL) is

Article History: Received 11 July 2018 ◆ Revised 20 August 2018 ◆ Accepted 22 August 2018

considered a successful model of teaching, helping both, to learn a foreign language and to acquire subject knowledge within that language (Marsh, Coyle, & Hood, 2010). At the same time, comparative studies (DESI-Konsortium, 2008; Prenzel et al., 2013) reveal heterogeneous knowledge among German students both, in reading and scientific literacy even if taught in their school language German. Hence, an educational CLIL programme for students without prior experience of CLIL needs to master heterogeneity, introduce CLIL methodology, facilitate subject-related learning and support multilingualism, assuring subject-related knowledge in both, the mother tongue and the foreign language (MSW NRW, 2012).

CLIL - Subject Learning in a Foreign Language

Content and language integrated learning (CLIL) denotes a teaching principle which underpins that content and language are taught in an integrated way (Marsh, Coyle, & Hood, 2010), whereas the German term "Bilingualer Sachfachunterricht" points to the primary focus on the subject content and competencies developed in two languages, with the foreign language to be predominantly used in CLIL lessons (MSW NRW, 2012).

Researchers like Cummins (1979, 1980) have been concerned with the development of bilingual learners. In his developmental interdependence hypothesis, Cummins (1979) assumes that a form of bilingualism both cognitively and academically beneficial can be achieved only based on an adequately developed first language skills. The fact that competencies gained in one language can subsequently be employed in both languages (Cummins, 1979, 1980) is explained with the common underlying proficiency (CUP) (Cummins 2007) which is held responsible for higher-order thinking. Moreover, the threshold hypothesis predicts that a maximum threshold level has to be accomplished to facilitate a bilingualism that is both, cognitively and academically beneficial (Cummins, 1979)**.**

In Germany, content and language integrated learning is either realised in streams with up to three subjects taught in English (MSW NRW, 2012) or in bilingual modules (Krechel, 2013) that may vary in length. For bilingual streams, Diehr (2012) has distinguished three types of CLIL classes as regards their employment and aim of foreign language and school language. As in type A only the foreign language is used, it aims at gains in foreign language proficiency and in subject-specific discourse competency. Type B is marked by a predominant use of the foreign language and employment of the school language especially in initial phases of CLIL methodology. In type C, both languages are used to gain content knowledge in both in order to develop true multilingualism (Diehr, 2012).

In literature, the positive influence of in-school CLIL interventions on foreign language competences is repeatedly emphasised (e.g. Bonnet, 2004; Bredenböker, 2000). Besides, it is argued that the authentic language use as a means of communication leads to better results, when recurrently used to this intent (Butzkamm, 1989). While the gain in language competency is often attributed to a functional language use and to a focus on meaning (van Patten, 1990), i.e. to a focus on working on the subject matter rather than on language learning, other authors (Heine, 2010) explain a language gain with a double effort on subject and language learning.

Cognitive Gain in the Scientific Subject Biology – The Development of Scientific Literacy

One main educational aim of the subject Biology is conceptual learning, the development of scientific literacy, and thus the development of cognitive academic language proficiency (CALP) (Cummins, 1980, 1999; Preisfeld, 2016; Richter & Zimmermann, 2003). To date, subject content learning has rarely been explored in scientific CLIL programmes (Bohn & Doff, 2010) and research studies document varied and contradicting results. On the one hand, many in-school investigations (e.g. Hartmannsgruber, 2014; Piesche, Jonkmann, Fiege, & Kessler, 2016) have revealed that students do not learn as successfully as in their school language. On the other hand, in and out of school, many authors (e.g. Buse & Preisfeld, 2016; Duske, 2017; Rodenhauser & Preisfeld, 2016, 2015) have identified equivalent competence gain in bilingual classes as compared to classes taught in their school language.

In natural sciences, practical experimentation is regarded to be the most important method to gain knowledge, conceptual understanding and develop problem-solving thinking (Engelen & Euler, 2004; Euler, 2009; Hammann, 2014; Hofstein & Lunetta, 1982; MSW NRW, 2014). Moreover, practical scientific out-of-school labs conducted in the school language have proved to offer favourable terms for the promotion of individuals (Damerau, 2012), working collaboratively, supporting each other in practical experimentation. Whether scientific lab days performed in a foreign language lead to homogeneous cognitive results, however, is rarely explored (Rodenhauser & Preisfeld, 2016, 2015). It has repeatedly been shown that the scientific content and methods of out-of-school activities should be integrated into the school syllabus (Berck & Graf, 2010).

Ability Self-concepts on Experimentation and Interpreting Data

The academic self-concepts have recurrently been identified as important predictors of academic achievements (Köller & Möller, 2010; Wolff et al., 2018). In research on science education, both the academic self-concept (Dickhäuser, 2006) and the experimentation competency (Schreiber et al., 2009) are investigated.

Hierarchically structured, the self-concept (Marsh, 1990) is a person´s perception of herself. As a part of it, the academic self-concept can be differentiated into a verbal and mathematic self-concept which can be divided into self-concepts of subject areas (e.g. Biology or English). The academic self-concept is formed through experiences with the learning environment and past performances influence subsequent behaviour. Hence, students externally assess their domain-specific abilities comparing their achievements to those of other students (social comparison). They may also assess them internally to their own prior achievements (temporal comparison) or to their achievements in other domains (dimensional comparison) (Wolff et al., 2018). Referring to the expectancy-value-theory of Eccles et al. (1983), the influence of experience of competency on the self-concept can be explained with a link of achievement performance, persistence and choice to expectancy-related as well as task-value beliefs (Eccles & Wigfield, 2002). Learners who have experienced successful performance of particular tasks, are more likely to choose demanding tasks in future. Applied to the out-of-school experimentation interventions, students may compare their current performance in the lab to their previous performances at school.

Some studies have shown that out-of-school labs with practical experimentation reinforce the students´ scientific self-concept (Euler, 2009). In some research (Damerau, 2012; Damerau & Preisfeld, 2016), the experimentation-related ability self-concept is investigated, which is considered a sub concept of the academic self-concept of the school subject Biology/ Sciences and itemised according to the three experimentation-related areas of hypothesising, conducting experiments and interpreting data, based on the model of experimentation competency by Schreiber et al. (2009).

While in this model (Schreiber et al., 2009) the area of planning includes hypothesising and making predictions, the area of conducting experiments comprises psychomotor skills like the handling of laboratory equipment, the setting up of experiments as well as the measurement and recording of results (Schreiber et al., 2009). In the area of interpreting data, both, the procession and interpretation of data as well as the testing of predictions and hypotheses mark cognitively and linguistically demanding tasks (MSW NRW, 2014), as these abilities interact with process-, product-related aspects (Schreiber et al., 2009) and communication skills**,** key issues addressed by curricula of both school subjects Biology and English (MSW NRW, 2014; MSW NRW, 2013).

Out-of-school labs doing practical experimentation have already documented increased short-term subject-related self-concepts (e.g. Brandt, 2005; Pawek, 2009) and experiment-related ability self-concepts (e.g. Damerau, 2012). Hence, it is to be explored, whether CLIL science programmes with a combination of two cognitively demanding factors develop students´ experimentation-related ability self-concepts equally well.

RESEARCH QUESTIONS

As empirical results on CLIL-effects on scientific content learning are contradictory, it remains questionable, whether CLIL interventions are equally beneficial in content learning, as compared to lessons taught in the school language. Moreover, it is worth knowing, whether the heterogeneity observed in regular classes leads to different cognitive gain depending on learner´s orientation.

RQ1. Does the combined educational concept of practical experimentation, biological content learning and CLIL lead to equally high cognitive achievement as a course conducted in the school language? Is cognitive gain dependent on learners´ orientation?

As there is still a lack of empirical research on the impact of CLIL interventions on the experiment-related ability self-concept, it is explored in this study.

RQ2. Does the educational concept realised impact the self-assessed ability self-concept on experimenting, irrespective of the working language used? Do students show different developments in this self-concept depending on their learner´s orientation?

RQ3. Does the educational concept realised impact the self-assessed ability self-concept on interpreting data, irrespective of the working language used? Do students show different developments in this self-concept depending on their learner´s orientation?

MATERIAL AND METHODS

Sampling and Setting

In a pre-post-follow-up-design, the cognitive and affective impact on senior students was measured at three reference times. Tests were performed at school one week prior, at the university campus directly after the lab day, and at school 4 to 6 weeks after the lab day in the presence of the investigator of this study. This quasi-experimental study was conducted between 2014 and 2016 with the experimental group taking part in an all-day CLIL course on neurobiology in the out-of-school lab performed in English as the working language. The control group attended an identical course at the university campus in their school language German. After the pre test, each group was given an experimental manual which contained both, the experimental procedure and the corresponding background information on neurobiology in English or German to prepare for the lab day at school, with additional word lists in English and German for the experiment group. The school teachers prepared the students for the lab day with the help of the experimental manual handed over by the investigator of this study. Of a total of 218 students present at all three reference times, 170 students constituted the CLIL experimental group being instructed in English and 48 students formed the control group being taught in German. The students´ mean age was 17.5 years, 62% of which were female. A majority of the 15 learner groups participating in this study (94 %) were educated in the two final years prior to Abitur (A-levels), one learning group deriving from the prior year. In the experimental group, 106 (64%) of the 170 students did not have prior contact with CLIL formats, whereas 36% of the students had. For the sample investigated in the present study (Buse, 2017), it could be shown that prior contact with CLIL methodology did not significantly influence the cognitive gain or the affective-motivational constructs explored in this study.

Educational Concept

As it is considered beneficial for its cognitive and affective impact that an out-of-school activity should be of relevance for school lessons (Berck & Graf, 2010), the contents and methods of the lab day are embedded in the school lessons, as part of preparation is done at school prior to the lab day and neurobiological aspects relevant for the lab day have been dealt with at school. Based on the concept of moderate constructivism (Reinmann & Mandl, 2006), an inquiry-based learning environment is created combining biological content with practical experimentation and the CLIL methodology. As the instructional elements, an experimental manual and the instructor mediate and coach the learners´ active, socially integrated and self-regulated knowledge construction. Content-wise and methodically, the lab day on neurobiology explores the phenomenon of the `learning brain´ on four levels with the use of different scientific methods and is thus linked to one of the four core contents of advanced Biology (MSW NRW, 2014): Cup stacking as an experiment on motor learning visualises behavioural changes. Exploring the brain´s activity in electroencephalography, comparing alpha-basic rhythm and its blockage, points to the imaging techniques and physiological dimension of the learning brain. Moreover, with the preparation of pig´s brain the anatomical level of learning and the spatial-functional relation of brain parts are examined. With the identification of pyramid cells, the microscopic and cellular level focuses on the type of neuron which is of major importance for learning (e.g. Bear et al., 2009). Hence, links are drawn between the four experimental sections of the laboratory day.

Students conduct a number of experiments, make observations and thus use and refine scientific procedures like microscopy already developed at school lessons. They also work with new kinds of methods, when they work with software on electroencephalography or prepare a pig´s brain. Moreover, initiated by focused tasks in the experimental script, students process the content and draw links between the results of the different experiments and the theoretical background. Experimentation is methodologically scaffolded with a lab report that follows the scientific method with question, hypotheses, investigation, conclusion to structure both, the collection and discussion of results and ideas. This lab report format is intended to help learners participate in a concise and focused subject-related communication in the working language and prepare them for a final discussion led by one work group at the end of the lab day. At the same time, the arrangement of group work of 3 to 4 learners is another important element of the lab day (Meyer, 2007), as it both requires and enhances collaboration when experimenting, discussing subject matter and using the respective working language. Prepared with the manual, students construct their knowledge during experimentation time and further refine it during the evaluation period. In their work groups (Vygotsky, 2002), learners compile their findings and interpretation of the experiments to finally give an oral presentation to their schoolmates.

Features of the CLIL Course

Both, the experimental and control group were taught identically with respect to subject-related content and methods. To meet the cognitive and linguistic challenges of the CLIL module (Krechel, 2013) of this study, the guiding principles of the lab day is “message before accuracy” (Long, 1991; Van Patten, 1990) to encourage students to freely voice their ideas on the scientific context in the foreign language and thus to both learn and communicate efficiently. Hence, as to the employment and aim of foreign and school language use in the experimental group, the educational CLIL programme can be classified as mixture of type A and B (Diehr, 2012). Although the foreign language is employed as the medium of learning to develop content knowledge (type A), the school language is used when needed (type B) and students are helped with English and German terms. Furthermore, the word list added to the experimental manual linguistically scaffold content learning and communication. In general, the all-day intervention performed in English, can be considered a foreign-language-intensive learning opportunity which is integrated into school lessons taught in the school language German. It can be assumed that some of the terms have already been taught in German in school lessons, whereas some of the technical terms and concepts were newly developed on the lab day. Thus, it may be assumed that the memory systems for cognitive academic language proficiency (CALP), existing distinct in school and foreign language, can be activated and expanded in this CLIL intervention (Diehr, 2016). Hence, linguistic and methodological scaffolds are meant to support students´ learning process and conceptual understanding in both, German and English.

INSTRUMENTS

Cognitive Test

The developed questionnaire included a cognitive test (Buse, 2017) arranged in 21 multiple-choice questions with multiple answers on the neurobiological content and the methods employed in the course, including four to eight items with one to four correct answers and a number of distractors. Identical questions were used at all three reference times in different order, the test language being German. Items were scored with one point for correct and no points for incorrect answers. A sample question is to be found in the Appendix (**Appendix 1**).

As the cognitive test consisted of content close to topics of the Biology curriculum (MSW NRW, 2014), its criterion validity was checked. Correlation of the test scores with the students´ last Biology grades yielded significant results ($r(217)= -.195$, $p\leq.01$), the negative correlation of which can be explained with the German grading scale from 1 for the best and 6 for the lowest grade.

Affective Questionnaire

In the affective questionnaire, the test language being German, two self-assessed experiment-related sub-competencies were investigated employing a five-point Likert-type scale from 0 for *strongly disagree* to 4 for *strongly agree*. Students´ ability self-concept on experimenting (**Table 1**) was measured using two items as in similar studies (Damerau, 2012) focusing on the self-assessed psychomotor abilities of handling laboratory equipment. The experimentation-related ability self-concept on interpreting data (**Table 1**) explored in this study (Buse, 2017) connects the self-assessed abilities of interpreting and communicating among peers and with experts. Thus, the items of this dimension derived from aims defined in the Biology curriculum (MSW NRW, 2014) concerning targeted communication and argumentation as well as constructive criticism, linking them with the working language employed in the intervention. The two experimentation-related ability self-concepts with a reliability of $.55\leq$Cronbachs $\alpha\leq .78$ were employed to explore the motivational-affective impact of the lab day on the learners.

Table 1. Experiment-related Self-concepts as Affective Variables with Scale Name including Number of Items, Reliability, Description and Sample Item

Scale (Number of items)	Reference time	Cronbach's α	Description	Sample item (original item)	Sample item (translation)
Experimenting (2)	Pre test	α = .61	self-concept on conducting experiments,	Im Umgang mit Laborgeräten bin ich gut.	I am good at handling laboratory equipment.
	Post test	α = .57			
	Follow-up test	α = .55			
Interpreting Data (8)	Pre test	α = .78	self-concept on communicating about interpreting data; connecting theory and practice;	Bezüge zwischen Versuchen oder Theorien und Versuchs-ergebnissen herzustellen, fällt mir leicht.	Creating references between different experiments or between theory and results is easy for me.
	Post test	α = .73			
	Follow up test	α = .72			

Table 2. Affective Variables with Scale Name and Items

Abbreviation	N=167	Item (original)	Item (translation)
Scientists	91	Ich bin eher naturwissenschaftlich interessiert und begabt.	I am primarily interested and talented in Sciences.
Language Lovers	41	Ich bin eher fremdsprachlich interessiert und begabt.	I am primarily interested and talented in foreign languages.
All-Rounders	35	Ich bin in beiden Bereichen gleich begabt.	I am talented in both areas.

To identify heterogeneous learner groups, as in similar studies (Rodenhauser, 2015), the students´ self-assessed preference (primarily scientifically interested, primarily interested in foreign language, interested in both) was collected in the affective questionnaire (See **Table 2** for scale names and items).

DATA COLLECTION

As it was not intended to pressurise the test subjects, the testing was carried out as a multiple-choice test of medium level of difficulty and adequate time for students to answer questions correctly was provided (Rost, 2004). The cognitive test and the questionnaire were employed in German for all groups at all reference times to prevent further language complications. The data collection was carried out under the supervision of the investigator both, in the classrooms and the premises of the out-of-school-labs.

DATA ANALYSIS

The analyses were carried out with SPSS (version 22 to 24). To gain a largely uniform sample for the cognitive test, error values were identified and substituted by means of the EM-algorithm (Expectation Maximization Algorithm) after determining the statistical significance values (Bortz & Döring, 2016). As the error values ranged from 0.5% to 9.7%, MCAR (Missing completely at random) was employed for replacement. A statically significant value of p=1.000 according to Little, measured with SPSS, proved that the missing data were missing at random.

As the aim of the cognitive test was to measure the cognitive performance, overly easy or overly difficult items were eliminated. Consequently, post test items with an index of difficulty between 10% and 90% in the post test were used in further steps of evaluation. After having excluded 4 items by the use of the item-difficulty-index, the discrimination index was determined for the remaining 108 items and all items with a discriminating power less than r<0.1 were eliminated, leading to 61 remaining items with a Cronbach´s alpha value of an internal consistency of Cronbach's α .66 (Buse, 2017). The reliability of the affective data was calculated in the same way. To allow comparison, the sum scores of the two experimentation-related self-concepts were normed to a maximum of 4.

To answer the drafted research questions on the changes in cognitive performance and experiment-related ability self-concepts, all data was first explored for pre test differences with independent samples t tests. With no *a priori* differences, repeated measures ANOVAs were employed to detect significant interaction group x

Table 3. Cognitive Test: Means and Standard Deviations for Pre, Post and Follow-up Test of Experimental and Control Group (Maximum=61 points)

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Experimental Group	32.59	3.76	38.63	5.30	36.61	4.77
Control Group	33.43	4.18	39.84	7.16	37.78	5.42

Table 4. Cognitive Achievement: Mean Values and Standard Deviations of Preference Subgroups for all three Reference Times (n=167): (Maximum=61 points)

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Scientists	32.69	3.71	38.44	5.12	36.43	4.48
Language Lovers	32.59	3.75	39.45	5.73	37.44	5.49
All-Rounders	32.57	3.89	38.52	5.18	36.38	4.57

reference time. Within the experimental group, dependent samples *t* tests were used to test for significant differences over time. In case of pre test differences, univariate ANOVAs with the post test or follow-up test sum scores as dependent variable and the pre test sum scores as covariate were conducted. In case there were more than 2 subgroups, post hoc tests according to Scheffé served to locate the identified differences in test scores. If identification of differences was impossible, dependent samples *t* test for the separate subgroups were performed.

RESULTS

Descriptive evaluation: In the experimental group, the subgrouping to the self-assessed preferences of the total 170 participants resulted in 54.5 % of the participants (N=91) classifying themselves as primarily biologically interested (*scientists*), whereas 24.6 % (N=41) of them saw themselves as primarily foreign-language-interested (*language lovers*) and 21 % (35 learners) rated themselves as interested in both (*all-rounders*) (**Table 2**). As three students did not reveal their preferences, the sample is reduced to 167 participants for these data.

Results of Cognitive Tests: For the two groups and the three reference times, the average scores of the cognitive test were measured. Corresponding means and standard deviations are shown in **Table 3.**

At the time of the pre test, an independent samples *t* test revealed no significant differences in test scores between the experimental and control group ($t(216)=1.372$, $p>.05$), i.e. the two groups showed a very similar cognitive achievement prior to the lab day. Additionally, repeated measures ANOVAs neither revealed statistically significant interactions between the three reference times and groups for the post test ($F(1,216)=.143$, $p=.706$) nor the follow-up test ($F(1.216)=.149$, $p=.700$) and nor between post and follow-up test ($F(1,216)=.003$, $p=.958$). These results indicate that the students in the experimental group both, equally well gained and retained knowledge (see **Table 3** for means and standard deviation).

As can be seen in **Table 3**, in the experimental group, the mean test scores on the cognitive achievement of the post test and follow-up test appear higher than those of the pre test. To test these differences for statistical significance, dependent samples *t* tests employed, revealed both, statistically significant differences in test scores from pre test to post test ($t(169)=12.96$, $p\leq .001$; $\eta^2= .4983$) and from pre test to follow-up test ($t(169)=9.959$, $p\leq .001$; $\eta^2= .3698$). For the experimental group, both significant short-term and long-term gains of cognitive knowledge were revealed.

Furthermore, potential differences in the gain of cognitive achievement over the three reference times were examined for the three preference subgroups. Means and standard deviation for the three groups at all reference times are given in **Table 4**.

A comparison between the three groups at the time of the pre test using a univariate ANOVA revealed no differences in cognitive knowledge prior to the lab day among the three preference groups ($F(2,164)= .019$, $p= .981$). Also repeated measures ANOVAs between pre test and post test ($F(2,164)= .542$, $p= .542$) and pre test and follow-up test ($F(2,164)= .722$; $p= .487$) did not show significant differences in average test scores. Independent of the assessed interests and talents, all students gained knowledge equally well.

Results of Ability Self-concept on Experimenting: A comparison of the mean values at time of the pre test revealed no significant differences between experimental group and control group ($F(1.216)= .205$; $p=$

Table 5. Self-concept Experimenting: Mean Values and Standard Deviations in Experimental and Control Group for all three Reference Times

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Experimental Group	2.73	0.73	2.83	0.72	2.74	0.73
Control Group	2.79	0.67	2.90	0.76	2.93	0.73

Table 6. Self-concept on Experimenting: Mean Values and Standard Deviations in Preference Groups at all the three Reference Times (n=167): (Maximum=4 points)

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Scientists	2.85	.68	2.95	.69	2.91	.72
Language Lovers	2.39	.70	2.50	.72	2.36	.63
All-Rounders	2.80	.80	2.93	.74	2.75	.72

.617). Neither between pre and post test ($F(1,216)= .002$; $p= .963$) nor between pre test and follow-up test ($F(1,216)=1.227$, $p= .269$) did a comparison of the mean values reveal significances between experimental and control group (see **Table 5** for mean values and standard deviation).

The analysis of overall longitudinal development within the experimental group, conducted with dependent samples t tests, yielded a significant difference in mean values between the pre test and post test ($t(169)=2.131$, $p= .035$, $\eta^2= .095$), but nothing significant from pre test to follow-up test ($t(169)= .421$, $p= .674$). Results documented a short-term, but no long-term increase in this kind of ability self-concept.

The ability self-concept on experimenting was further tested for potential differences in the preference subgroups over the three reference times (see **Table 6** for mean values and standard deviation).

At time of the pre test, a univariate ANOVA revealed significant differences among the three preference subgroups ($F(2,164)=6.287$, $p\leq .002$, $\eta^2= .071$). As can be seen in **Table 6**, students with scientific preferences *(scientists)* as well as *all-rounders* showed a high scientific self-concept, whereas students with preferences for foreign languages *(language lovers)* showed a low self-concept. The comparisons using Scheffé post hoc criterion of significance indicated that prior to the treatment, significant differences were registered between the pairs: scientists and language lovers ($p < .01$), all-rounders and language lovers ($p < .05$) and scientists and all-rounders ($p > .05$).

A comparison from pre to post test between the preference groups employing a univariate ANOVA, in which the difference in the self-concept at the time of the pre test ($F(2,164)=6.287$, $p\leq .002$, $\eta^2= .071$) was used as a covariate, conveyed no differences ($F(3,163)=1.341$, $p= .264$). However, a comparison from pre to follow-up test, using the pre test differences as a covariate, revealed significant differences in the self-concept developed ($F(3,163)=1.341$; $p< .05$, $\eta^2= .042$). Dependent samples t tests employed for each preference group and the three reference times did not yield any significant difference. When looking at the follow up test results (**Table 6**), however, there seems to be a tendency towards a long-term increase for the scientists, while the long-term scientific self-concept of both, the language lovers and the all-rounders is back to pre test level.

These results on the ability self-concept on experimenting indicate that a significant short-term increase in the scientific self-concept, independent of preferences, could be observed in the experimental group, similar to the control group. Thus, a scientific CLIL course in an out-of-school lab appears to be equally beneficial for all three learner groups. Furthermore, a tendency towards a long-term increase can be observed for scientists.

Results of the Ability Self-concept on Interpreting Data: As can be seen in **Table 7**, at the time of the pre test, mean values of the self-concept on interpreting data were significantly higher in the experimental group as compared to the control group ($t(1,216)= 2.494$; $p\leq.05$ $\omega^2= .023$). These pre test results considered as covariates, the following univariate ANOVAs revealed a statistically significant effect of reference time for the post test for the two groups ($F(1,215)= 4.777$, $p\leq .05$, $\eta^2= .022$), but no significant effect at the time of follow-up test ($F(1,215)=3.807$, $p= .052$) (see **Table 7** for mean values and standard deviations). These results indicate that the mean average of the self-concept on interpreting data developed differently in the short-term. While it decreased in the treatment, it remained constant in the control group. Long-term it developed similarly.

Table 7. Self-Concept on Interpreting Data: Mean Values and Standard Deviations in Experimental Group and Control Group at three Reference Times (Maximum 4points)

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Experimental Group	2.25	.54	2.13	.53	2.27	.50
Control Group	2.02	.69	2.16	.66	2.01	.53

Table 8. Self-Concept on Interpreting Data**:** Mean Values and Standard Deviations in Preference Groups at all the three Reference Times (n=167): (Maximum=4 points)

Group	Pre Test		Post Test		Follow-up Test	
	M	SD	M	SD	M	SD
Scientists	2.31	.50	2.15	.46	2.33	.51
Language Lovers	2.13	.57	2.12	.57	2.14	.49
All-Rounders	2.45	.58	2.12	.67	2.30	.50

To test this ability self-concept among the preference subgroups for potential differences over the three reference times repeated univariate ANOVAs were employed (see **Table 8** for mean values and standard deviation), but did not reveal any significant differences at pre test time (F(2,164)= 1.711, p= .184), at time of post test (F(2,164)= .054, p= .948) or at the time of the follow-up test (F(2,164)= 2.215, p= .112). In the experimental group, these results indicate a short-term decrease that develops independently of the preference voiced by learners.

DISCUSSION

Cognitive Evaluation

The research question on the cognitive gain attained in a practical experimentation situation in a foreign language can be answered positively. It appears that subject-related proficiency developed independently of the working language, both, short- and long-term, irrespective of preference groups. With respect to its homogeneous increase on subject-related competences in the scientific out-of-school experimentation programme in CLIL format, the results of this study are in line with some in- and out-of-school studies (Duske, 2017; Rodenhauser & Preisfeld, 2015). However, they disagree with other CLIL studies in school (e.g. Hartmannsgruber, 2015; Piesche et al., 2016).

It may be reasonable to assume that the homogeneous cognitive gain documented in this study is to be explained multi-causal. As to Cummins´ hypotheses (1979), students benefit from CLIL interventions, if they have exceeded a threshold level in linguistic competency in the foreign language. It may be surmised that, in general, the tasks of the intervention met a medium level of linguistic and cognitive competency, as the average English language proficiency of advanced students (MSW NRW, 2013) may have helped to meet even language-wise and cognitive demands of tasks (Thuermann 2013). The relevance of contents for school (Berck & Graf 2010) is often considered a favourable factor for subject learning. Furthermore, it is often suggested that the approach of "focus on message" (Long, 1991; van Patten, 1990) is believed to support cognitive gain in CLIL programmes. The cognitive achievement in the foreign language is often explained with a deeper level of processing (Craik & Lockhart, 1972). Moreover, social contact is thought to support learning (Vygoskij, 2002), when students discuss with peers and experts on a subject matter in a foreign language and make use of offered scaffolding if needed, to meet task expectations and thus actively construct their knowledge.

Evaluation on Self-Concepts

The first group of research questions focused on the influence of the present intervention on the ability self-concept on experimenting, i.e. it explored the relevance of the working language on it. Furthermore, it was the research question, whether this intervention influenced learners according to their preferences (students primarily interested in sciences, primarily interested in languages, interested in both). The results of this study did not verify an impact of the working language on the development of this ability self-concept, a fact which may be linked with the experimentation competency being considered a primarily psychomotor skill. The documented short-term gain in the ability self-concept on experimenting can be attributed to the fact that the self-concepts is formed by experience, due to an internal comparison of the current experience to prior experience (Wolff et al., 2018), for instance. Thus, the gain in self-concept can be explained with experience of competency (Deci & Ryan, 1985) in learners during the practical laboratory work, whereas the long-term disappearance of this effect may be attributed to lack of opportunity to practical experimentation in regular school lessons (Graf & Berck, 2010). The observed tendency of a long-lasting strengthened self-concept in the students assessing themselves as primarily interested in sciences, can be explained with the positive impact of a remembered successful experience, either due to internal or external comparison (Wolff et al., 2018). Hence, the study shows that for students interested in sciences an intervention combining practical experimentation with foreign language use may be advantageous to promoting their ability self-concept.

The second group of research questions was, whether the intervention at hand impacted the ability self-concept on interpreting data and, thus, whether its development was dependent on the working language. The research question also implied, whether this ability self-concept was developed differently for the self-assessed preference groups. With a short-term increase of the self-concept in the control group and a short-term decrease in the experimental group, a dependence of the working language was documented in this study. The short-term decreased ability self-concept with students being taught in English may be attributed to external and internal comparison. However, it may also be explained with a more realistic assessment of their abilities as to the demands on the laboratory day (Engeln, 2004). Although a combined use of practical experimentation and English as the working language is regarded as a beneficial opportunity of acquiring language proficiency (Bohn, 2012), it may also be true that enjoyment and motivation are reduced when students meet (initial) challenges of adapting to CLIL methodology (Marsh, Coyle, & Hood, 2010). Moreover, it may be argued that for the highly complex competency of interpreting data the students have felt the discrepancy of cognitive and linguistic abilities in the CLIL intervention of this study (Thürmann, 2013).

Conclusion

In this study, students who participated in an out-of- school laboratory day in CLIL format attained subject knowledge equally as high as students educated in the school language. Moreover, knowledge acquisition appeared homogeneously, independent of learner´s preferences. It may be concluded that scientific experimentation interventions can be successfully employed with CLIL to the benefit of heterogeneous learner groups. Further studies may explore factors that impact subject-related learning and conceptual understanding in scientific CLIL programmes. Hence, it may be of interest to investigate different forms of CLIL lessons regarding the use of foreign and school language (Diehr, 2012). Moreover, further studies may explore the gain of subject knowledge in both, the foreign and the school language with quantitative and qualitative research.

We could also show that a CLIL format of a science lab day can strengthen the ability self-concept on experimenting. Consequently, these results underpin that the combined use of practical experimentation and language appeared to be beneficial for the development of subject-related self-concepts.

However, this study also revealed significant differences in the development of the ability self-concept on interpreting data between the experimental and control group, with lower mean scores for the experimental group being taught in English. These results can either be attributed to a more realistic assessment or to the high-order thinking task and a discrepancy of cognitive and linguistic competences. Studies may further explore the area of interpreting data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Margarethe Buse – University of Wuppertal, Department of Zoology and Biology Education, Germany.

Karsten Damerau – University of Wuppertal, Department of Zoology and Biology Education, Germany.

Angelika Preisfeld – University of Wuppertal, Department of Zoology and Biology Education, Germany.

REFERENCES

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2009). *Neurowissenschaften. Ein grundlegendes Lehrbuch für Biologie, Medizin und Psychologie*. [Neurosciences – A textbook for Biology, Medicine and Psychology]. Heidelberg: Spektrum Akademischer Verlag. https://doi.org/10.1007/978-3-8274-2228-6

Berck, K.-H., & Graf, D. (2010). *Biologiedidaktik. Grundlagen und Methoden*. [Biology Education]. Wiebelsheim: Quelle & Meyer.

Bohn, M. (2013). Biologie [CLIL Biology]. In: W. Hallet & F. G. Königs (Eds.), *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning*. [Handbook of bilingual education. Content and language integrated learning] (pp.286-295). Seelze: Friedrich Verlag.

Bohn, M., & Doff, S. (2010). Biologie bilingual: Die Perspektive der Unterrichtspraxis. In: S. Doff (Ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung*. [Bilingual subjects in secondary education. An introduction] (pp. 72-88). Tübingen: Narr Francke Attempto Verlag.

Bonnet, A. (2004). *Chemie im bilingualen Unterricht*. Kompetenzerwerb durch Interaktion. [Chemistry in bilingual education]. Opladen: Leske & Budrich. https://doi.org/10.1007/978-3-663-09491-3

Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. [Research methods and evaluation in social sciences and humanities]. Heidelberg, Berlin: Springer Verlag.

Brandt, A. (2005). *Förderung von Motivation und Interesse durch außerschulische Experimentierlabors*. [Promoting motivation and interest through out-of-school labs]. Göttingen: Cuvillier.

Bredenbröker, W. (2000). *Förderung der fremdsprachlichen Kompetenz durch bilingualen Unterricht. Empirische Untersuchungen*. [Promoting foreign language competencies through bilingual education]. Frankfurt a. M., New York: Peter Lang

Buse, M. (2017). *Bilinguale englische experimentelle Lehr-Lernarrangements im Fach Biologie. Konzeption, Durchführung und Evaluation der kognitiven und affektiven Wirksamkeit*. [Bilingual teaching in Biology. Its cognitive and affective-motivational impact]. Dissertationsschrift. Bergische Universität Wuppertal.

Buse, M., & Preisfeld, A. (2016). The Emperor Moth – The Power to Develop. Sachfachliches Lernen im bilingualen Biologieunterricht der Sekundarstufe I. [The Emperor Moth – The power to develop. Subject-related learning in bilingual Biology lessons] (pp. 201 –219). In B. Diehr, A. Preisfeld, & L. Schmelter (Eds.), *Bilingualen Unterricht erforschen und weiterentwickeln*. Frankfurt a. M., Bern: Peter Lang.

Butzkamm, W. (1989). *Psycholinguistik des Fremdsprachenunterrichts. Natürliche Künstlichkeit. Von der Muttersprache zur Fremdsprache*. [Psycholinguistics of foreign language teaching]. Tübingen: Francke.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL. Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Craik, F. I. M., & Lockhart, R.S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behaviour, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research, 49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (1980). The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue. *TESOL Quarterly, 14*(2), 175–187. https://doi.org/10.2307/3586312

Cummins, J. (1999). *BICS and CALP: Clarifying the Distinction*.

Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics, 10*(2), 221-40.

Damerau, K. (2012). *Molekulare und Zell-Biologie im Schülerlabor. Fachliche Optimierung und Evaluation der Wirksamkeit im BeLL Bio* (Bergisches Lehr-Lern-Labor Biologie). [Molecular and cell biology in an out-of-school laboratory]. Dissertationsschrift.

Damerau, K., & Preisfeld, A. (2016). Einfluss von Schülerlaborbesuchen auf das Fachwissen und das Fähigkeitsselbstkonzept heterogener Schülergruppen. [The influence of out-of-school lab days on cognitive gain and self-concepts in heterogeneous learner groups]. In U. Gebhard & M. Hammann (Eds.), *Bildung durch Biologieunterricht. Internationale Tagung der Fachsektion Didaktik der Biologie im VBIO, Hamburg 2015* (pp. 337–353). Innsbruck, Wien, Bozen: VBIO.

Deci, E, L., & Ryan, Richard M. (1985). *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum Press. https://doi.org/10.1007/978-1-4899-2271-7

DESI-Konsortium (Ed.) (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie. Ergebnisse der DESI-Studie*. [Education and competency gain in German and English]. Weinheim: Beltz.

Dickhäuser, O. (2006). Fähigkeitsselbstkonzepte. Entstehung, Auswirkung, Förderung. [Self-concepts: formation, impact, promotion]. *Zeitschrift für Pädagogische Psychologie, 20*(1-2), 5-8. https://doi.org/10.1024/1010-0652/a000084

Diehr, B. (2012). What´s in a name? Terminologische, typologische und programmatische Überlegungen zum Verhältnis der Sprachen im Bilingualen Unterricht. [What´s in a name? Reflections on terminology, typology and programme regarding the relation of languages in bilingual education]. In B. Diehr & L. Schmelter (Eds.), *Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven* (pp.17-36). Frankfurt a.M.: Peter Lang.

Diehr, B. (2016). Doppelte Fachliteralität im bilingualen Unterricht. Theoretische Modelle für Forschung und Praxis. [Development of literacy in both languages in bilingual education]. In B. Diehr, A. Preisfeld, & L. Schmelter (Eds.), *Bilingualen Unterricht erforschen und weiterentwickeln* (pp.57-84). Frankfurt a.M.: Peter Lang.

Duske, P. (2017). *Bilingualer Unterricht im Fokus der Biologiedidaktik. Auswirkungen von Unterrichtssprache und -kontext auf Motivation und Wissenserwerb*. [Bilingual education in the focus of Biology didactics]. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-16492-8

Eccles, J. S., Adler, T. F., Futtermann, R., Goff, S. G., Kaszala, C. M., Meece, J. L., & Migdely, C. (1983). Expectancies, values and academic behavior. In J. T. Spence (Ed.), *Achievement and achievement motives*. San Francisco: W.H. Freeman.

Eccles, J., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. Annu. Rev.Psychol., 53, 109-32. https://doi.org/10.1146/annurev.psych.53.100901.135153

Engeln, K. (2004): *Schülerlabors. Authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*. [Out-of-school laboratories: Arousing interest in sciences]. Dissertation. Kiel.

Engeln, K., & Euler, M. (2004). Forschen statt Pauken. Aktives Lernen im Schülerlabor. [Researching instead of cramming: Active learning in an out-of-school laboratory] *Physik Journal, 3*(11), 45–48.

Euler, M. (2009). 25 Schülerlabore: Lernen durch Forschen und Entwickeln. [25 out-of-school laboratories: Learning by researching and developing] (pp.799-818). In E. Girdwidz, R. Häußler, & P Kircher (Eds.), *Physikdidaktik. Theorie und Praxis*. Heidelberg: Springer Verlag.

European Commission (Ed.) (2004). *Promoting language learning and linguistic diversity: An action plan 2004-06*. Office for Official Publications of the European Communities. Retrieved from http://www.saaic.sk/eu-label/doc/2004-06_en.pdf

Gnutzmann, C., Jackisch, J., & Rabe,F. (2015). Deutsch und Englisch als Wissenschaftssprachen. Eine Interviewstudie mit Wissenschaftlern, Herausgebern und Verlagsmitarbeitern. [German and English - the languages of science]. *Fremdsprachliches Lehren und Lernen, 44*(1). 9-28.

Hammann, M. (2014). Experimentieren. [Conducting experiments]. In U. Spörhase-Eichmann & W. Ruppert (Eds.). *Biologie Methodik. Handbuch für die Sekundarstufe I und II* (pp.102–106). Berlin: Cornelsen Verlag.

Hartmannsgruber, M. (2014). *Bilinguale Biologie. Konzeption und Evaluation*. [Bilingual Biology. Development and evaluation]. Baltmannsweiler: Schneider Hohengehren.

Heine, L. (2010). *Problem solving in a foreign language. Studies on language acquisition, Vol. 41*. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110224467

Hofstein, A., & Lunetta, V. N. (1982). The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Review of Educational Research, 52*(2), 201–217. https://doi.org/10.3102/00346543052002201

Köller, O., Möller, J. (2010). Selbstwirksamkeit. [Self-efficacy]. In D. H. Rost (Ed.), *Handwörterbuch pädagogische Psychologie*. Weinheim: Beltz.

Krechel, H.-L. (2013). Organisationsformen und Modelle in weiterführenden Schulen. [Organisation and models of bilingual education]. In W. Hallet & F. G. Königs (Eds.), *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (pp.74–80). Seelze: Friedrich Verlag.

Long, M. H. (1991). Focus on form: a design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. J. Kramsch (Eds.), *Foreign language research in cross-cultural perspective. Studies in bilingualism, 2*, 39–52. Amsterdam, Philadelphia: J. Benjamins Pub. Co. https://doi.org/10.1075/sibil.2.07lon

Marsh, H.W. (1990). The Structure of Academic Self-Concept: The Marsh/Shavelson Model. *Journal of Educational Psychology, 82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Mayer, J. (2014). Forschendes Lernen. [Explorative Learning]. In U. Spörhase-Eichmann & W. Ruppert (Eds.), *Biologie Methodik. Handbuch für die Sekundarstufe I und II* (pp. 107–113). Berlin: Cornelsen.

Meyer, H. (2007). *Unterrichtsmethoden*. [Teaching methods]. Frankfurt a. M.: Cornelsen Scriptor.

MSW NRW (2014). *Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen Biologie*. [Curriculum of advanced Biology in North Rhine-Westphalia]. Retrieved from http://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/bi/KLP_GOSt_Biologie.pdf

MSW NRW (2012). *Bilingualer Unterricht. Biologie deutsch-englisch in der Sekundarstufe I*. [Bilingual education in Biology in North Rhine-Westphalia in forms 5 to 10]. Retrieved from http://www.schulentwicklung.nrw.de/cms/upload/bilingualer_Unterricht/documents/HR_BU_BioE_SekI_0912.pdf

MSW NRW (2013). *Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen Englisch*. [Curriculum of advanced English in North Rhine-Westphalia]. Retrieved from https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/e/GOSt_Englisch_Endfassung3.pdf

Pawek, C. (2009). *Schülerlabor als interessefördernde außerschulische Lernumgebung für Schülerinnen und Schüler aus der Mittel- und Oberstufe*. [Out-of-school laboratories as interest-arousing learning environment for students of secondary education]. Kiel.

Piesche, N., Jonkmann, K., Fiege, C., & Keßler, J.-U. (2016). CLIL for all? A randomised controlled field experiment with sixthgrade students on the effects of content and language integrated science learning. *Learning and Instruction, 44*, 108–116. https://doi.org/10.1016/j.learninstruc.2016.04.001

Preisfeld, A. (2016). Die Bedeutung bilingualen Experimentalunterrichts in Biologie für die fachliche und sprachliche Kompetenz. [The importance of CLIL and practical experimentation in Biology for the development of scientific literacy and linguistic competencies]. In B. Diehr, A. Preisfeld, & L. Schmelter (Eds.), *Bilingualen Unterricht erforschen und weiterentwickeln* (pp.103-23). Frankfurt a. M: Peter Lang.

Prenzel, M., Sälzer, C., Klieme, E., & Köller O. (Eds.) (2013). *PISA 2012 – Fortschritte und Herausforderungen in Deutschland*. [PISA 2012 – progress and challenges in Germany] Münster: Waxmann.

Reinmann, G., & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten. [Teaching and instructional design] In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie. Ein Lehrbuch* (pp. 613–658). Weinheim, Basel: Beltz.

Richter, R., & Zimmermann, M. (2003). Und es geht doch: Naturwissenschaftlicher Unterricht auf Englisch. [Bilingual education in natural sciences] In M. Wildhage & E. Otten (Eds.), *Praxis des bilingualen Unterrichts* (pp. 116–146). Berlin: Cornelsen-Scriptor.

Rodenhauser, A. (2016). *Bilinguale biologische Schülerlaborkurse. Konzeption und Durchführung sowie Evaluation der kognitiven und affektiven Wirksamkeit*. [Bilingual out-of-school laboratory courses in Biology. Development, implementation and evaluation of cognitive and affective influence]. Dissertation.

Rodenhauser, A., & Preisfeld, A. (2015). Bilingual (German–English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. *International Journal of Environmental and Science Education, 10*(1), 99-10.

Rodenhauser, A., & Preisfeld, A. (2016). A Glue from Snail Slime?! – Umsetzung und Evaluation eines bilingualen Moduls für den Biologieunterricht. In B. Diehr, A. Preisfeld, & L. Schmelter (Eds.), *Bilingualen Unterricht erforschen und weiterentwickeln* (pp. 221-43). Frankfurt a. M: Peter Lang.

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion*. [Textbook on test theory and test construction]. Bern: H. Huber

Schreiber, N., Theyßen, H., & Schecker, H. (2009). Experimentelle Kompetenz messen?! [Evaluating experimental competencies]. *Physik und Didaktik in Schule und Hochschule, 8*(3), 92–101.

Thürmann, E. (2013). Spezifische Methoden für den Bilingualen Unterricht/CLIL [Specific methods of bilingual education]. In W. Hallet & F. G. Königs (Eds.), *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (pp. 229–235). Seelze: Friedrich Verlag.

Van Patten, B. (1990). Attending to Form and Content in the Input. *Stud. Sec. Lang. Acq., 12*(3), 287–301. https://doi.org/10.1017/S0272263100009177

Vygotsky, L. S. (2002). *Denken und Sprechen. Psychologische Untersuchungen*. [Thinking and speech]. Weinheim. Basel: Beltz.

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the Effects of Social, Temporal, and Dimensional Comparisons on Academic Self-Concept. *Journal of Educational Psychology*, 1-21. https://doi.org/10.1037/edu0000248

APPENDIX

Appendix 1 Sample question of cognitive test

Original question with corresponding items

Bei selbst hergestellten Färbepräparaten von Pyramidenzellen

- ☐ sind Soma, Zellkern, Dendriten und Axone gut sichtbar und unterscheidbar.
- ☐ sind Soma und Zellkern sichtbar, Dendriten und Axone im Ansatz sichtbar, jedoch nicht unterscheidbar.
- ☐ sind Soma, Zellkern und Axon sichtbar.
- ☐ sind besonders die Dendriten und Axone sichtbar.

Translation

In the microscopic preparation manufactued by us

- ☐ *soma, nucleus, dendrites and axons are to be seen well and distinguishable.*
- ☐ *soma and nucleus are visible, parts of dendrites and axons are visible, but cannot be distinguished*
- ☐ *soma, nucleus, dendrites and axons are visible.*
- ☐ *dendrites and axons can well be distinguished.*