

# Measuring the Knowledge Increase of Eight Grade Students in a Bilingual Biology Unit

*Midiendo el incremento del aprendizaje de estudiantes de octavo grado en una unidad de biología bilingüe*

*Medindo o aumento do conhecimento de estudantes do 8º ano em uma unidade bilíngue de biologia*

189

---

**Stephanie OHLBERGER**

[orcid.org/0000-0003-3354-420X](https://orcid.org/0000-0003-3354-420X).

Bielefeld University, Germany.

[stephanie.ohlberger@uni-bielefeld.de](mailto:stephanie.ohlberger@uni-bielefeld.de)

---

**Claas WEGNER**

[orcid.org/0000-0001-8866-7704](https://orcid.org/0000-0001-8866-7704).

Bielefeld University, Germany.

[claas.wegner@uni-bielefeld.de](mailto:claas.wegner@uni-bielefeld.de)

---

Received: 2017-12-04

Send to for peer review: 2018-01-19

Accepted by peers: 2018-01-19

Approved: 2018-01-22

---

To reference this article in APA style / Para citar este artículo en APA / Para citar este artigo  
Ohlberger, S., & Wegner, C. (2017). Measuring the knowledge increase of eight grade students in a bilingual biology unit. *Latin American Journal of Content and Language Integrated Learning*, 10(2), 189-216. doi:10.5294/lacil.2017.10.2.2

**ABSTRACT.** In recent years, the issue of using bilingual education has come to the public's attention. A widespread belief is that students in bilingual classes achieve a higher competence in the language, but that they lag behind in subject-specific knowledge when compared to fellow students in traditional (i.e. monolingual) classroom settings. Nevertheless, by evaluating knowledge gain in a short teaching unit, previous studies have shown that bilingual students rarely experience drawbacks. Although there are a variety of different opinions, the present study aims at detecting how subject-specific knowledge is influenced by bilingual biology lessons. The study was conducted at a grammar school in North Rhine-Westphalia, Germany, with 13- and 14-year-old students. To assess differences in students' knowledge gain, test and control groups were used, in which students were taught 12 identical biology lessons in a bilingual and a traditional classroom. The unit consisted of basic characteristics seen in (social) insects. Both classes took a knowledge test before and after the unit. We hypothesised that there would be a higher increase in knowledge for the bilingual class compared to the monolingual class. Results suggested a slightly higher knowledge gain in bilingual students, but no significant differences were revealed.

**Keywords:** CLIL; bilingual education; biology; knowledge increase; Germany.

**RESUMEN.** En los últimos años, la educación bilingüe ha atraído la atención del público. Una creencia común es que, aunque los estudiantes desarrollan su competencia en lengua en un alto grado cuando asisten a clases bilingües, su aprendizaje del contenido específico de las asignaturas es menor si se compara con el de los estudiantes que asisten a clases en un ambiente tradicional (monolingüe). Sin embargo, al evaluar el conocimiento adquirido durante una unidad corta, se ha demostrado que los estudiantes bilingües rara vez experimentan dicha desventaja. Aunque hay variedad de opiniones, el presente estudio pretende descubrir la influencia que pueden tener las clases bilingües de biología sobre el conocimiento específico de dicha asignatura. El estudio se llevó a cabo en un colegio de primaria en la ciudad alemana de North Rhine-Westphalia, con estudiantes de entre 13 y 14 años. Para evaluar las diferencias relativas al conocimiento de los estudiantes, se empleó un grupo de control y uno experimental, en los que los estudiantes de ambos grupos recibieron 12 sesiones de biología idénticas en un salón de clase bilingüe y en uno tradicional. En la unidad, se estudiaron las características básicas observadas en insectos (sociales). Ambos grupos presentaron un examen de conocimientos antes y después de la unidad. Nuestra hipótesis consistía en que podría haber un aumento mayor de conocimiento en la clase bilingüe comparado con la clase monolingüe. Los resultados sugirieron un aumento ligeramente superior en los estudiantes bilingües, aunque no se observaron diferencias significativas.

**Palabras clave:** AICLE; educación bilingüe; biología; mejoramiento del aprendizaje; Alemania.

**RESUMO.** Recentemente, a questão do uso da educação bilingue chamou à atenção do público. Uma crença difundida é que estudantes de aulas bilingues alcançam uma competência mais alta no idioma, mas apresentam atraso no conhecimento de temas específicos quando comparados a colegas em configurações de salas de aula tradicionais (isto é, monolíngues). No entanto, ao avaliar o ganho de conhecimento em uma unidade curta de aprendizagem, estudos anteriores mostraram que estudantes bilingues raramente experienciam obstáculos. Embora existam muitas opiniões diferentes, o presente estudo busca detectar como o conhecimento de um tema específico é influenciado por aulas bilingues de biologia. O estudo foi conduzido na escola primária North Rhine-Westphalia, na Alemanha, com estudantes de 13 e 14 anos. Para avaliar as diferenças no ganho de conhecimento dos estudantes foram usados grupos de teste e controle nos quais os estudantes receberam 12 aulas de biologia idênticas em salas de aula bilingues e tradicionais. A unidade consistia de características básicas encontradas em insetos (sociais). Ambas as turmas realizaram um teste de conhecimentos antes e depois da unidade. Nossa hipótese é de que haveria maior aumento no conhecimento para a turma bilingue comparado à turma monolíngue. Os resultados sugerem um aumento de conhecimento ligeiramente maior em estudantes bilingues, mas não foram encontradas diferenças significativas.

**Palavras-chave:** AICL/CLIL; Alemanha; aumento do conhecimento; biologia; educação bilingue.

## Introduction

---

The study at hand researches how bilingual classes influence subject-specific knowledge in biology. From a parent's perspective, the decision becomes difficult when faced with choosing to enrol their child in a bilingual or a mainstream class. Apart from additional language benefits (Bredenbröker, 2000; Dalton-Puffer, 2008; Lasagabaster, 2008), it seems reasonable that parents might fear that their child will lag behind in that particular subject, as he/she must not only keep up with the content but deal with it in another language. Furthermore, teachers are not entirely convinced about the effectiveness of bilingual education as well (Müller-Schneck, 2006), and it is considered particularly difficult to assess students in bilingual programmes, as it remains uncertain if mistakes are caused by the language barrier or by lack of knowledge.

The article will introduce the concept and organisation of bilingual education in Germany and give a short overview of current research about content learning. Not only will we explore the relevance of bilingual education from a student's perspective, but problems and fears that parents might have about bilingual education will also be addressed. Subsequently, a research question and hypotheses will be raised, which are followed by a description of the study's design. Finally, the study's results will be presented and discussed, ending with concluding remarks to encourage further research.

## Theoretical Background

### Bilingual education in Germany

Bilingual education has developed from a unique phenomenon to a mainstream movement (Appel, 2011, p. 85). The concept's promotion by politics, the Ministry of Education and several cultural affairs committees lead to its inclusion in many school profiles (Appel, 2011, p. 85). Generally speaking, bilingual classes can be structured in various ways, causing further confusion by the use of different terms and implementations. It should be noted that this article uses CLIL, bilingual education/classes/lessons, and other terms as synonyms. Further,

the conceptual design of bilingual education in Germany, with an emphasis on the federal state of North Rhine-Westphalia, is focused on.

Content and Language Integrated Learning (CLIL) is an accepted teaching approach for bilingual education throughout Europe. It suggests that an additional language is not only used for learning, but for teaching both content and language as well. Content and language are tightly linked, even though the emphasis may shift between the two from time to time. Thus, CLIL can be seen as an innovative form of language and subject teaching, which is closely related to bilingual education and immersion (Coyle, Hood, & Marsh, 2010, p. 1).

Bilingual classes are commonly offered at grammar schools and aimed at middle-class students. Earlier accusations stated that only elite students were offered the opportunity to participate in bilingual education. Nowadays, the recent introduction of bilingual classes in all kinds of primary and secondary schools counters that argument. There are roughly two groups of CLIL students. In one category, there are students who learn parts of their subject in a foreign language (usually English and French, sometimes Spanish). In the second, there are students who speak German as a second language and complete their entire school career in German. The latter case applies to a noticeably higher number of students, for example immigrants, but is not structured as thoroughly as common CLIL classes. This issue gained major importance since 2015, when many refugees came to Germany and when schools decided to integrate these students into so-called “international” or “Welcome” classes in order for them to learn the language of teaching first (for an overview of German practices, see Schmiedebach & Wegner, 2017).

Generally, CLIL is regarded as an additive late partial immersion, as it tends to start in year seven with one subject, and more subjects are added or exchanged later (Breidbach & Viebrock, 2012, p. 6).

In our study’s experimental school, history is taught bilingually in year 7, and is substituted for geography and biology in year 8; all three of these subjects make up the bilingual education profile in year 9. In every year, there is a maximum of two bilingual classes and, in order for the students to take part, they must have at least average marks in English (Evangelisch Stiftisches Gymnasium, 2015). This organisation is known as bilingual branches or streams, although there is the possi-

bility to offer bilingual modules in individual subjects. In this case, only particular topics are taught in a foreign language over a certain period of time (Breidbach & Viebrock, 2012, p. 6).

Students in bilingual branches usually receive two additional English lessons per week. This sums up to a total of seven lessons, compared to the five lessons their fellow students have. This allows for the enhancement of general-language, subject-specific, and methodological vocabulary (Müller-Schneck, 2006, p. 84).

### **Current state of research**

In order to contrast this study in the context of other types of research on subject-specific knowledge changes, this section will focus on studies that follow the same principles with regard to study focus and design. It will be comprised almost entirely of studies conducted in German schools and will look at the issues of content learning in bilingual programmes and performance assessment.

Koch and Bündler (2008) did not find validation for the most common objection that subject-matter learning is inhibited by the use of a foreign language. Supporting findings were consistently confirmed by individuals, but rarely empirically backed up (Koch & Bündler, 2008, p. 4). Therefore, Koch and Bündler (2008) investigated four natural scientific project groups in year five: a German-speaking group, an English-speaking group without experience and two groups that already had bilingual experience from primary school. The materials were adjusted so that language supporting elements were integrated into the lessons. This helped the teacher maintain the foreign language during the whole period, even though monolingualism was not enforced. Knowledge gain on the content level was evaluated by concept mapping, which is a way of illustrating knowledge structures. The highest increase for content knowledge was recorded for the two test groups with bilingual experience, and the bilingual groups also had a higher learning success than the German control group (Koch & Bündler, 2008, pp. 5–7). Language difficulties may have caused slower communication, but it appears that, as a consequence thereof, the students had a higher attention span and consolidation of knowledge. Without exception, the language problems resulted in a sensitisation and reflection of the learning processes. The authors recommended designing the lessons

as action-oriented and communicative, along with the use of written impulses to promote independent work (Koch & Bänder, 2008, p. 7).

In his well-known DEZIBEL project, Zydatiſ (2007) surveyed 180 sixteen-year-old students in bilingual branches of grammar schools in Berlin (Zydatiſ, 2007, p. 72). He wanted to investigate correlations between linguistic competence (English) and subject-matter abilities in geography, history, and biology. Since his study was carried out at experimental schools, there were no consistent syllabi and the subject knowledge analysis had to focus on “interdisciplinary text and discourse competences.” Listening and reading comprehension, text-type bound writing, and the overall use of English was analysed using a proficiency test and an achievement test (Zydatiſ, 2007, p. 83). A quarter of the sample took an oral test on language competence. To examine the subject-specific text and discourse competence, a 130-minute test was used which included reading and interpreting non-fictional texts, understanding and applying technical terms, and writing so a certain text-type could be assessed. Background data such as socioeconomic status, Anglophone cultural affinity, and subjective state in the class was collected (Zydatiſ, 2007, p. 73). A comparison was drawn between bilingual students that took tests in English and students in the regular programme that had to answer the same tasks in German. Results were consistently positive; in all three competence-based exams bilingual students scored just as good as or higher than the control group. As expected, their competence level for the English language in listening comprehension, speaking, vocabulary knowledge, and the correct use of written language was distinctly higher when compared to the control group. Regarding subject-specific text and discourse competence, the performance of bilingual and regular students was mostly equal, with some significant differences in favour of the bilingual students, even though to a clearly lower extent than for the language competences (Zydatiſ, 2007, p. 320).

A similar study on the students’ scientific literacy revealed that bilingual students in biology achieved a distinctly higher competence increase than their monolingual peers (Osterhage, 2007). Scientific literacy is the educational objective of science teaching and therefore reflects the students’ subject-specific knowledge. Osterhage (2007) examined nine ninth grade classes in three different grammar schools

in Bremen through a 45-minute test. The test contained 18 questions in the areas of “reproduction and sexuality,” “respiration and photosynthesis,” and “biochemistry and nutrition”. All test questions were taken from PISA 2003, an international science test that assesses students’ scientific literacy. The author found that bilingual students outperformed their monolingual peers in five competence areas: convergent thinking, number processing, graph processing, mental models, and verbalising facts. He concluded that subject-matter learning is not inhibited in bilingual classes but that other situational circumstances should be considered, such as small classes and additional biology lessons (Osterhage, 2007, pp. 43–47).

When it comes to measuring the performance of bilingual students, Coyle, Hood, and Marsh (2010) suggest three methods for evaluating an increase in content level knowledge. According to the authors, translating a test into the mother tongue is not a reasonable way to collect reliable and comparable data, as tasks requiring detailed answers might be too challenging (Coyle, Hood, & Marsh, 2010, p. 138). A student’s understanding should be expressed by other means through the use of graphics, tables, or oral presentations (Coyle, Hood, & Marsh, 2010, p. 138). They also view tests in the mother tongue as a legitimate way of examining students’ knowledge, while also admitting that special terms might only have been learned in the CLIL language (Coyle, Hood, & Marsh, 2010, p. 139). A third evaluation method is a portfolio of work, which enables the assessment of different levels of understanding (Coyle, Hood, & Marsh, 2010, p. 139). In our study, we decided to use an English translation of the test; data is relatively comparable as multiple-choice tasks were used and extensive answer texts were not required. In addition, it was decided not to have the test for the bilingual class in their native language because the experimental school’s teaching principles require using the target language as much as possible.

Kondring and Ewig’s (2005) study also addressed the issue of performance assessment in bilingual classes. In general, evaluating students’ performance in biology is already difficult, and only gets tougher if it is a bilingual class (Kondring & Ewig, 2005, p. 50). Therefore, the researchers considered whether a similar performance could be expected for bilingual and mainstream students, and in which language the assessments should be carried out (Kondring & Ewig, 2005, p. 50). They investigated

two ninth grade classes at a grammar school in North Rhine-Westphalia in a small-scale study. Both classes received an identical teaching unit on the topic 'hearing' by the same teacher. They were assessed via a 20-minute test with mostly closed-form questions (labelling and allocating) and two directional open questions. Only students of the same performance level in both groups were compared; this was established using a record of science and language grades for each student (Kondring & Ewig, 2005, p. 51). An additional survey found that 35–40% of bilingual teachers frequently used multiple-choice questions, which is rarely done in classes only taught in German (Kondring & Ewig, 2005, p. 55). It also appeared that teachers always correct language mistakes in written tests in bilingual courses, even though these mistakes do not influence the assessment (Kondring & Ewig, 2005, p. 56). At least a quarter of the teachers in bilingual classes commonly exchanged views on grading their students (Kondring & Ewig, 2005, p. 58). Results did not show any significant differences in subject-matter performance between bilingual and mainstream students. Instead, there was a slight tendency for regular students to excel the bilingual students performance-wise, as these students supposedly mastered technical terms slightly better (Kondring & Ewig, 2005, p. 59). However, the small sample size does not allow for a representative status (Kondring & Ewig, 2005, p. 58). Our study uses a structure like Kondring and Ewig's study; there is an identical teaching unit for a test and control class, and the type of school and testing procedures were the same.

Similarities can also be found in a study done by Haagen-Schützenhöfer, Mathelitsch, and Hopf (2011) using a bigger sample size. Since there are nearly no scientifically proven assertions about subject-matter competence, the researchers evaluated bilingual physics classes with regard to the linguistic and content knowledge acquisition of students in the eleventh class. With the help of a pre- and post-test, the learning results of a bilingually taught test group ( $n = 127$ ) and a control group taught in their native language ( $n = 78$ ) were compared. Results demonstrated no apparent differences in content performance of all students. The lack of a significant difference argues against the common opinion that a foreign language impedes subject-specific learning success. However, it should be noted that the teaching unit



only consisted of four consecutive physics lessons (Haagen-Schützenhöfer, Mathelitsch, & Hopf, 2011, p. 238ff.).

In addition, Dalton-Puffer (2008) found that CLIL students are able to keep up with monolingually taught peers in subject-matter competence (p. 4). She emphasises that a foreign language has the capability to improve the learning of subject-specific concepts. This is closely related to Bonnet's study, which implies that switching to the mother tongue does not solve conceptual problems in the foreign language, but dealing with a concept in L2 can allow for deep semantic processing (as cited in Dalton-Puffer, 2008, p. 142). All the same, performance loss in some students should not be overlooked (Washburn; Nyholm, as cited in Dalton-Puffer, 2008, p. 142).

Lamsfuß-Schenk's (2008) study found that ninth grade students taught in a different language were better acquainted with the content knowledge in history than a control group taught in their native language, which may be attributed to dealing more intensively with the learning content (*deep processing*) (Wolff, 2011, p. 80). Another important factor was an increase in motivation and interest in the topic, leading Wolff to conclude that bilingual education gives the subject new impetus by increasing students' motivation (Wolff, 2011, p. 80).

### **Problems and fears**

A common objection arises when using foreign languages in class, namely that one cannot learn as much as they could in their native language since mastering the subject matter now will take considerably more time. On the other hand, classes getting through the content much faster are not bound to have lasting effects either (Bonnet, Breidbach, & Hallet, 2009, p. 188).

The most common fear reported in parents is that the content level for each lesson will be lowered. Teachers who have not taught bilingual classes yet also share this view (Müller-Schneck, 2006, p. 262). It has been reported that the necessity of language simplification would also entail subject-matter degradation (Müller-Schneck, 2006, p. 262; Hajer, as cited in Dalton-Puffer, 2008, p. 142). Likewise, language problems influence general teaching style, since there is often the need to resort to teacher-centred learning (Pilz, as cited in Müller-Schneck, 2006, p. 262).

In most cases, alleged difficulties can be avoided by using didactic-methodological strategies (Müller-Schneck, 2006, p. 263), e.g., word inference strategies, repeating commonly used phrases for describing and explaining, and also allowing students enough time to gather their thoughts, to talk to their neighbour etc. However, more suitable material (including word annotations, easy vocabulary and alike) must be made available, and teachers should receive special training (Müller-Schneck, 2006, p. 264).

This is well demonstrated by studies that have shown that teachers lacking the ability to phrase questions in a way that can address students' individual language performance levels is at least partly responsible for poor subject-matter performance (Yassin, Tek, Alimon, Baharom, & Ying, 2010, p. 52). This creates a dilemma as teachers lower the level of difficulty for both content and language (Yassin et al., 2010, p. 52). One particular study used closed questions without any discussion, and this led to no development of the subject matter (Yassin et al., 2010, p. 52).

From the students' perspective, the benefits of bilingual education are again often experienced differently. A graduate survey of German-French students showed that even in hindsight, 72% thought they had learned at least as much as their monolingually taught peers had. Only 2% admit that they did not learn as much (Landesinstitut Soest, 1995; also see Christ and Krechel, as cited in Müller-Schneck, 2006, p. 270).

Another criticism in respect to bilingual classes is that only linguistically talented students benefit from the concept, but that those who perform well in the topic-related subject and poorer in the language do not (Appel, 2011, p. 85). Another problem is the increasing selection of high-performance students with their parents eager for an optimal academic career (Bonnet, Breidbach, & Hallet, 2009, p. 194). Unfortunately, this observation has to be confirmed in the study at hand. The school used for the study generally appeals to a rather privileged student body, which is why mostly well-educated students choose the bilingual branch.

Even though some studies (Washburn and Nyholm, as cited in Dalton-Puffer, 2008, p. 142) argue against bilingual education's success, one should be careful about making broad generalisations. This study

aims to take a look at the current situation in North Rhine-Westphalia to discover if the contrary is possible. As known, situational conditions and success factors must be taken into account before ultimately deciding on a position, which is too positive or negative (Appel, 2011, p. 87).

## Method

### Research questions

This study deals with knowledge acquisition in a biology unit taught in German for a control group and in English for a test group. Both groups consisted of year eight students, and the teaching unit focused on the topic of insects. The motivation to conduct this investigation was the feeling that parents' and teachers' constantly express fear that students would not learn in the same depth and speed in bilingual classes as their monolingually taught peers. In contrast to the studies reported earlier (particularly Kondring & Ewig, 2005 and Haagen-Schützenhöfer, Mathelitsch, & Hopf, 2011), this study was spread over 12 lesson-teaching units. This gives us reason to suggest that significantly positive results may be achieved with a longer testing period. Thus, our overarching research question becomes: Is knowledge acquisition in the subject of biology higher in a bilingual class when compared to a regular class?

The research question will be reviewed with the following hypotheses. Initially, it must be guaranteed that students have learned about insects during the teaching unit. For this purpose, there are two hypotheses:

*Hypothesis 1a: The control class taught in German shows a significant performance difference between the points of measurement  $t_1$  and  $t_2$ .*

*Hypothesis 1b: The test class taught in English shows a significant performance difference between the points of measurement  $t_1$  and  $t_2$ .*

After the assessment of knowledge acquisition in both classes, differences between the groups will be analysed. As previous studies occasionally found positive effects for bilingual classes, the following hypothesis is presented:

*Hypothesis 2: Knowledge acquisition in the test class is significantly higher than in the control class.*

Appel (2011) emphasises that only linguistically talented students will benefit from bilingual classes (Appel, 2011, p. 85). Even though the claim for individual learning support also holds true for bilingual education, students with a gift for languages certainly have an advantage. This circumstance will be analysed with the third hypothesis:

*Hypothesis 3: Students in the test class will show a positive correlation between English performance and knowledge acquisition.*

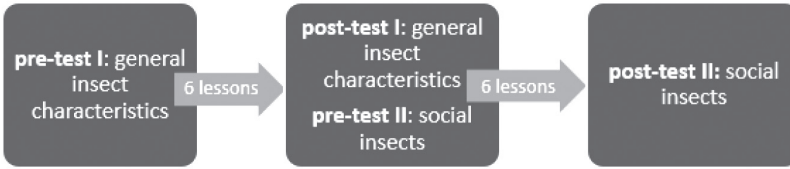
In general, the study targets the analysis of subject-matter acquisition only, since language improvements have been frequently confirmed in previous studies. In addition, these results could help draw conclusions about the suitability of biology as a bilingual subject.

### **Study design**

Fifty students were taken from a year eight in a grammar school in NRW. Both the test group and control group ( $n = 25$  each) had 12 girls and 13 boys. Throughout the teaching unit, students occasionally missed a lesson, resulting in a final analysis of 44 tests. The test group received their tests in English, whereas the control group took their tests in German; the tests were identical because the English test was a direct translation of the original test. The test group received a vocabulary list that was displayed for everyone on the board. At the end of each test, all students were asked to fill in a short self-evaluation in German.

To determine the suitability of the knowledge test, the questions were given as a pilot test in an eighth grade class that was not part of the actual study. The test was based on 12 lessons that were carried out in both classes. In order for the students not to feel overwhelmed, the usual test duration of 20 minutes was kept by splitting the test into two halves (see Figure 1). Apart from this division for practical reasons, the knowledge test will be regarded as a coherent test instrument.

**Figure 1.** Study set-up with the points in time of respective knowledge tests

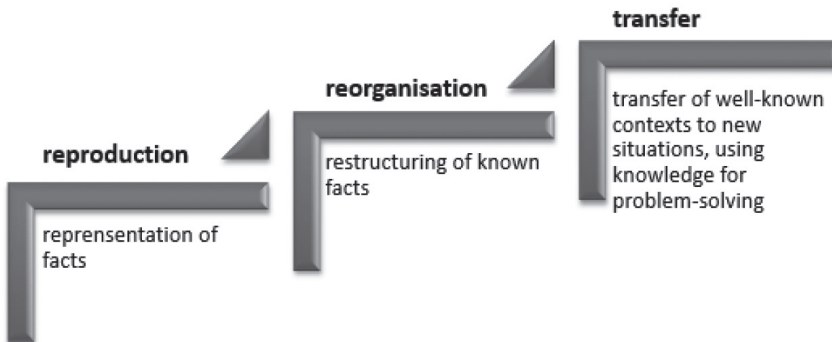


Source: Own elaboration.

### Knowledge test

The written performance review in this study had to meet certain criteria; the questions must represent different requirement areas, meaning the level of difficulty. This would allow students to receive additional points for the more complex and demanding tasks in higher requirement areas (Berck, 2005, p. 123). Three requirement areas have been established by the Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in Germany and comprise *reproduction*, *reorganisation*, and *transfer/problem-solving* (Spörhase, 2012, p. 294; see Figure 2). Questions formulated to fit these requirements should touch upon the four competence areas of *subject-specific knowledge*, *knowledge acquisition*, *evaluation* and *communication* (Kultusministerkonferenz, 2004). Whether a task corresponds to requirement area I, II or III always depends on the previous instruction (Spörhase, 2012, p. 295).

**Figure 2.** Requirement areas as stated in German curricula



Source: Own elaboration, based on Spörhase (2012, p. 294f).

Due to the special test situation for the study in the school, most tasks were allotted to requirement area I (reproduction), with nearly three times as many tasks in this category compared to reorganisation tasks. The test instrument neglected requirement area III because students had just begun their bilingual biology classes and were not as used to bilingual tests as older students.

There was an imbalance in the thematic distribution of tasks, because the sub units covered different amounts of material; thus, topics that were dealt with lengthily were covered by tasks in both requirement areas, whereas ambiguous topics were made up of only reproduction tasks. A few sample questions (see Table 1) illustrate the type of test and the kind of questions that were asked.

**Table 1.** Sample questions from the knowledge test on (social) insects

Question	Requirement area / Level of difficulty	Type of question
Which of these animals belong to the class of insects? a. Beetles, centipedes, spiders, bees b. Flies, butterflies, isopods, millipedes c. Bees, beetles, butterflies, ants d. Isopods, millipedes, grasshoppers, wasps	I	Multiple choice
How is the event called when the workers drive away the drones? a. drone battle b. nuptial flight c. metamorphosis d. moving out	I	Multiple choice
Label the insect leg.	I	Semi-open → labelling
Give two examples of adaptations of a common insect's walking leg. Name the respective animal that has developed a different leg and describe the changes.	II	Open → short answer
A food source is 700 m away from the beehive, at a 40° angle to the left from the position of the sun. Draw the dancing pattern.	II	Open → drawing

Source: Own elaboration.

The self-evaluation at the end of every test was answered with the help of a ranking scale. Students were allowed to answer the evaluation after the regular test was completed. It was acceptable for all students to give feedback in German, as meta-performance is already challenging in one's mother tongue.

The internal consistency of the knowledge test was measured by Cronbach's  $\alpha$  at  $t_2$ , including test results of 44 students. Unfortunately, the 26-question test achieved only a Cronbach's  $\alpha$  of 0.584. In order to test for reliability, the questions were also examined under the criteria of discriminatory power and difficulty index. The discriminatory power is calculated for each item individually and should be as high as possible, preferably over 0.5 (Bortz & Döring, 2006, p. 220). Students with a high overall score in the test will also have high scores in distinct items, whereas it is the other way round with students scoring low (Bortz & Döring, 2006, p. 219).

Usually, only a few students will answer difficult questions correctly, while there are many right answers in the case of easier questions (Bortz & Döring, 2006, p. 218). This circumstance is considered by the difficulty index; if the influence of very difficult or very easy questions becomes too great, the test does not give any indication about students' performance and differences between students anymore (Bortz & Döring, 2006, p. 219). In the best case, a test includes questions of different levels of difficulty, ranging from values of 0.2 to 0.8 (Bortz & Döring, 2006, p. 219).

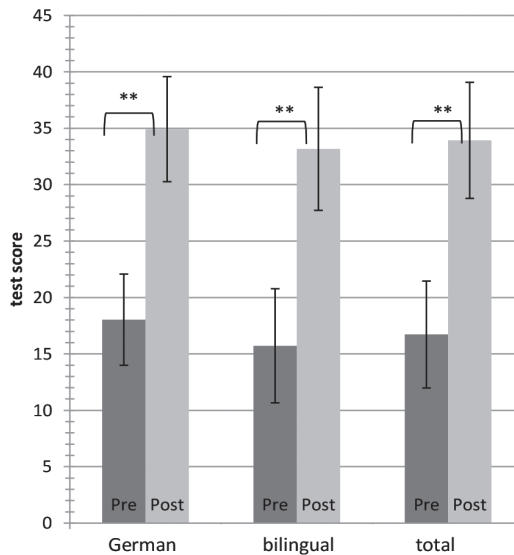
The discriminatory power as well as the difficulty index for each item can be seen in Appendix I. By including the test results of the pilot students, a sample size of 72 students was reached. Due to too low or high difficulty indices, items 8, 24, 25, 26, 33, and 34 cannot be included in the test evaluation. Items 2, 8, 10, 17, 20, 24, 25, 27, 32, and 34 do not have the prescribed discriminatory power. It was to be expected that the open questions (8, 11, 14, 24, and 34) would not be answerable without any pre-knowledge, which is why questions 8, 24, and 34 were still kept in the test as a good indicator of knowledge increase. Also, question 26 remains, as it is one of the most difficult ones amongst questions stemming predominantly from the requirement areas I and II. In total, eight multiple-choice questions had to be removed from the test due to calculations of difficulty and discriminatory power, resulting in a total of 48 points to be achieved.

## Results

To compare classes, biology grades from the school term between 2014/2015 were analysed. The German grading system is based on a scale from 1 to 6, with 1 being the best mark. The bilingual test class ( $n = 25$ , mean = 2.28) had slightly better marks than the control class ( $n = 19$ , mean = 2.47).

When looking at mean scores for the pre- and post-tests, the control class had slightly better results (see Figure 3). The difference accounts for 3.0 points in the pre-test (mean<sub>(German)</sub> = 18.03, mean<sub>(bilingual)</sub> = 15.72) and approximately 1.5 points in the post-test (mean<sub>(German)</sub> = 34.92, mean<sub>(bilingual)</sub> = 33.18). It thus begins to show that the difference turns out only half as much in the post-test compared to the pre-test. On average, the score of both classes taken together was 16.72 for the pre-test and 33.93 in the post-test (SD  $\pm$  4 and 5 points).

**Figure 3.** Means of test scores in pre- and post-test.\*



\* The means of total test scores in pre- and post-test for the German ( $n=19$ ), bilingual ( $n=25$ ) and overall ( $n=44$ ) group. SD(German pre)  $\pm$  4,04, SD(German post)  $\pm$  4,66, SD(bilingual pre)  $\pm$  5,05, SD(bilingual post)  $\pm$  5,45, SD(overall pre)  $\pm$  4,74, SD(overall post)  $\pm$  5,14. \*\*  $p < 0.01$ .

Source: Own elaboration.



Calculation of the performance difference between  $t_1$  and  $t_2$  is based on the score difference between the two tests. Accordingly, the knowledge gain in the German control class was 16.89 points (SD  $\pm$  4.17) and 17.46 points (SD  $\pm$  5.52) in the bilingual test class (see Figure 3). As seen in Figure 3, the differences between the pre- and post-tests are significant for both classes ( $p < .001$ ) (see tables 2 and 3), verifying hypotheses 1a and 1b.

**Table 2.** t-Test for differences in test results in the bilingual test class between the pre- and post-test (n = 25)

	Mean	SD	t	Df	Sig. (2-tailed)
Score difference Pre – Post	17.45600	5.52450	15.799	24	.000

Source: Own elaboration.

**Table 3.** t-Test for differences in test results in the control class between the pre- and post-test (n = 19)

	Mean	SD	t	Df	Sig. (2-tailed)
Score difference Pre – Post	16.88947	4.16998	17.655	18	.000

Source: Own elaboration.

The second hypothesis was concerned with the students' learning success, therefore we analysed whether the test group's knowledge gain was higher than in the control class. As previously mentioned, bilingual students outperformed the control class by 0.57 points in terms of knowledge gain (see Figure 3). However, this result is insignificant, allowing us to discard the second hypothesis.

The third hypothesis suggested that there was a positive correlation between the students' English mark and knowledge acquisition in the bilingual class. The correlation calculation (see Table 4) acknowledges a weak negative but insignificant correlation ( $r = -.325$ ). The negative correlation values come about by the German marking system, which puts the smallest number (1) as the best mark a student can get, while a higher number (up to 6) is worse.

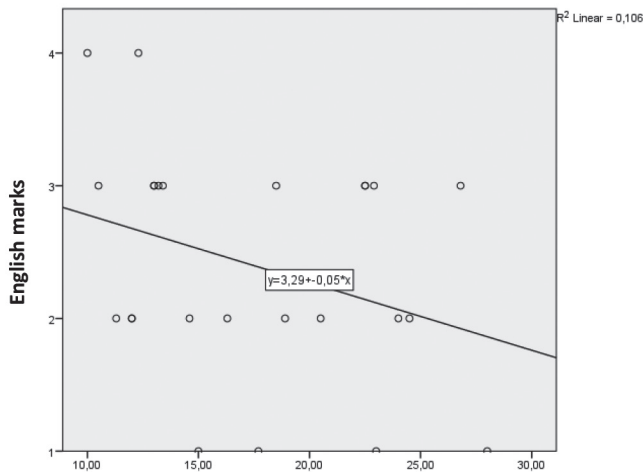
**Table 4.** Correlation between English mark and knowledge acquisition for bilingual students (n = 25)

		Knowledge gain in points	English mark
Knowledge gain in points	Pearson correlation	1	-.325
	Sig. (2-tailed)		.112
	N	25	25
English mark	Pearson correlation	-.325	1
	Sig. (2-tailed)	.112	
	N	25	25

Source: Own elaboration.

Students with good English marks (1 or 2 on their mid-year report) exhibited an average knowledge gain of 18.29 points, while the weaker students (3 or 4 on their mid-year report) yielded an average knowledge gain of 16.55 points. Although the difference between mean values is insignificant, there was still a tendency for linguistically talented students to experience a slightly higher knowledge gain than their fellow students (see Figure 4).

**Figure 4.** Graphic display of correlation between the English mark and knowledge acquisition for bilingual students (n=25)



Source: Own elaboration.

In general for all 44 students, a significant negative correlation ( $r = .314$  and  $p = .038$ ) between the biology mark and the students' knowledge gain was seen (see Table 5 and Figure 5).

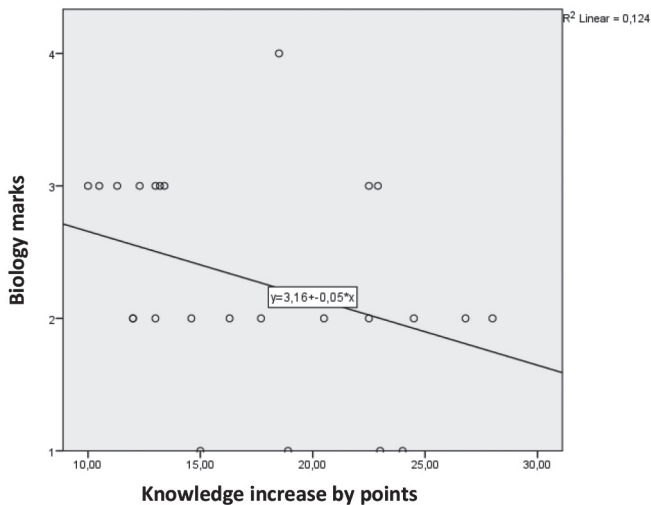
**Table 5.** Correlation between the biology mark and knowledge acquisition for all 44 students ( $n = 44$ )

		Biology mark	Knowledge gain in points
<b>Biology mark</b>	Pearson correlation	1	-.314*
	Sig. (2-tailed)		.038
	N	44	44
<b>Knowledge gain in points</b>	Pearson correlation	-.314*	1
	Sig. (2-tailed)	.038	
	N	44	44

\* The correlation is significant at a level of 0.05 (two tailed).

Source: Own elaboration.

**Figure 5.** Graphic display of correlation between the biology mark and knowledge acquisition for all students ( $n=44$ )



Source: Own elaboration.

When separating the groups, the correlation is no longer significant (see tables 6 and 7). The comparison between the biology mark and knowledge gain in the German-speaking control group turns out to be quite weak ( $r = -.242$ ), whereas the bilingual test group shows at least a moderate correlation ( $r = -.352$ ).

**Table 6.** Correlation between the biology mark and knowledge acquisition for the control group (n = 19)

		Biology mark	Knowledge gain in points
<b>Biology mark</b>	Pearson correlation	1	-.242
	Sig. (2-tailed)		.318
	N	19	19
<b>Knowledge gain in points</b>	Pearson correlation	-.242	1
	Sig. (2-tailed)	.318	
	N	19	19

Source: Own elaboration.

**Table 7.** Correlation between Biology mark and knowledge acquisition for the test group (n = 25)

		Biology mark	Knowledge gain in points
<b>Biology mark</b>	Pearson correlation	1	-.352
	Sig. (2-tailed)		.084
	N	25	25
<b>Knowledge gain in points</b>	Pearson correlation	-.352	1
	Sig. (2-tailed)	.084	
	N	25	25

Source: Own elaboration.

If we only take the bilingual test group into consideration, it becomes apparent that the correlation of the biology mark and the students' knowledge increase ( $r = -.352$ ) is slightly higher than the one between the English mark and knowledge increase ( $r = -.325$ ), but both are still insignificant.

## Discussion

---

Significant differences between the pre- and post-test demonstrate that students in both classes learned the material, justifying H1a and H1b. The groups do not differ as much in the post-test when compared to the pre-test, suggesting a very high knowledge gain in the test group, possibly attributed to a greater commitment or possibly a higher capability. One factor to consider is an initial struggle with comprehension in the pre-test, which may result in lower scores. This may also explain the difference in mean scores between the pre- and post-test.

The maximum knowledge gain was achieved by a bilingual student. Learning success performance was 5% higher in the test group when compared to the control group. Despite the fact that a total score of 48 points was not reached, this may be due to multiple reasons: a) the open answer tasks were too challenging for the students; b) students were not used to the item format; or c) students did not answer because of convenience (no time left, they had to prioritise).

In general, the study's multiple-choice test format was oriented towards similar study designs (Kondring & Ewig, 2005; Müller-Schneck, 2006; Haagen-Schützenhöfer, Mathelitsch, & Hopf, 2011). It was assumed that the bilingual class was more familiar with this question type than the control class, which might have influenced performance. Apart from a few vocabulary questions in the bilingual class, no questions were asked about the test tasks.

The second hypothesis was discarded, as the knowledge gain in the bilingual class was not significantly higher than in the control class. Even though the bilingual teaching of the subject matter could not prove a more intense promotion of the test students, the teaching unit was equally effective in both classes. It did not appear

that bilingual students were at a disadvantage in terms of learning the subject matter. These findings coincide with other studies that found that bilingual students did not perform weaker in any of the competence-oriented tests than their fellow students taught in German (Zydati, 2007; Kondring & Ewig, 2005; Haagen-Schtzenhfer, Mathe-litsch, & Hopf, 2011). These authors interpreted their results as proof that a foreign language does not impede learning success.

Results concerning the second hypothesis could be reinforced by the students' self-evaluation sheets. All students claimed that the post-test was easier to complete, implying that answering the ques-tions was easier, thus some learning process was involved. In addition, the control class considered the test easier, which is presumably due to the fact that the foreign language did not play a role in their situation. The bilingual students perceived the pre-test as "very difficult" and the post-test as "appropriate," whereas the students taught in German con-sidered the pre-test as only "fairly difficult" and the post-test as "rather easy." The different perception about difficulty is most certainly due to the test group's test being in English and thus including more words they did not know before. Even though many students suspected that the post- and pre-test test would be the same, comments following the post-test advocated content with their answers and that the level of difficulty was adequate.

The third hypothesis proposed a positive correlation between En-glish marks and knowledge acquisition, but had to be falsified due to insignificant results. A moderately negative correlation ( $r = -.325$ ; due to Germany's marking system, see results) between the students' En-glish mark and knowledge acquisition was observed. The correlation is insignificant, implying that content and pre-knowledge in biology may be more important than profound language skills. However, if chosen to look at the biology mark instead, a different picture emerges. The difference of approximately one point between the groups may be interpreted when it is assumed that only high-performance students are considered to form a bilingual class. Within the bilingual class, the stronger students experienced a significantly higher knowledge acqui-sition than the weaker ones. This might be explained because stronger

students only have to understand the content in a new language once they are relatively confident with the subject-specific content. Weaker students must overcome two obstacles—they need to pay attention to both language *and* content. This complies with the idea that students who do well in the subject but poorly in the language should be able to profit from bilingual education (Appel, 2011, p. 85). Our results show three cases in which students with a high subject performance (biology mark in the previous term was 2) were weaker than their language performance (English mark in the previous term was 3). Even so, this did not reflect a lower knowledge acquisition. It ensures that language is not as big of an obstacle for the understanding of subject content as previously assumed. Ultimately, there were only a relatively few amount of students that did not do well in English.

It must be taken into consideration that the specific students with high biology marks may have a higher motivation for the subject itself. This could possibly minimise language difficulties. This assumption is firmly supported by authors who deem that bilingual classes yield their highest success rate with an increase in motivation and interest in the topic (Wolff, 2011, p. 80; Coyle, Hood, & Marsh, 2010, p. 140). In addition, Lasagabaster (2008, p. 39) admitted that the results of his study might be influenced because CLIL students were generally more talented and motivated to learn in English.

Also, students' individual perspectives about the test are relevant for the third research hypothesis. Bilingual students considered the post-test as "rather easy" language-wise compared to the pre-test. The content, however, was graded as more difficult than the language in both tests. The self-evaluation also discussed whether students would feel more confident when being able to answer test questions in German, but they declared a switch to German to be unnecessary, with a slight preference for having the questions formulated in German. This gives the impression that students are eager to answer the questions in English. Positive attitudes seen in the post-test show that, after acquiring basic vocabulary and content, there is no preference for answering questions in the mother tongue. These findings support the issue that a foreign language is not problematic if it is used for teaching subject content.

## Conclusion and outlook

---

Our aim was to investigate the extent to which subject-specific learning would be inhibited by a foreign language—a fear commonly expressed in parents and teachers. It was concluded that both the test and control class exhibited a significant performance difference before and after the teaching unit, implying a knowledge gain. Although the increase was slightly higher in the bilingual class, there was no significant difference in comparison with the control class.

A moderately positive correlation was found between English marks and knowledge gain in bilingual students, although it appeared that their performance in biology was considerably more crucial. Bilingual students good in biology achieved a significantly higher knowledge gain compared to their weaker classmates. This increase in knowledge was still seen in the linguistically weaker student. This may be explained by the fact that the language deficit is compensated by higher levels of motivation and enthusiasm about the subject.

Future studies should include an additional teacher questionnaire to record individual teaching differences (Osterhage, 2007). This is particularly important if lessons are not taught by the same teacher. To expand upon this, the regular teacher should give bilingual classes alone, as an external teacher may be inclined to falsify the intervention by using extra time to settle in and introduce the situation to the students. The class composition must also be taken into consideration, as there may already exist performance differences between bilingual and monolingual students.

Even though one should not jump to the conclusion that teaching bilingual classes is the *ne plus ultra*, its opportunities should still be recognised and further developed. Bilingual education will continue to be used more extensively in the future, therefore it becomes particularly important to search for improvements. In recent years, investigations have shown various independent methodologies for bilingual classes (Bach, 2001, p. 65). Didactic and methodological developments support an existing established teaching model, but now matters of systematised educational programmes and pedagogical goals become important to keep in mind (Bach, 2001, p. 67). Study designs will also



need to adjust as current evaluations about the effectiveness of bilingual classes are usually performed on a small scale through single classes or schools (Bach, 2001, p. 66) and an unintentional allocation of test and control groups (Breidbach & Viebrock, 2012, p. 14). Large sample sizes are hard to come by, since confounding variables such as varying levels of performance and class size remain. Therefore, it should be considered to focus on individual classes, (qualified) teaching staff and the content that is learned by CLIL students as well as motivational and emotional aspects involved in CLIL teaching (Breidbach & Viebrock, 2012, p. 14).

## Acknowledgements

---

This project is part of the *Qualitätsoffensive Lehrerbildung*, a joint initiative of the Federal Government, and the Länder, which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research. The authors are responsible for the content of this publication.

## References

---

- Appel, J. (2011). Two for the price of one? Leise Zweifel am bilingualen Sachfachunterricht. *Forum Sprache - Die Online-Zeitschrift für Fremdsprachenforschung und Fremdsprachenunterricht*, 3(2), 85–88.
- Bach, G. (2001). Bilingualer Unterricht. *Englisch*, 36(2), 64–67.
- Berck, K.-H. (2005). *Biologiedidaktik – Grundlagen und Methoden* (3rd Ed.). Wiebelsheim: Quelle & Meyer.
- Bonnet, A., Breidbach, S., & Hallet, W. (2009). Fremdsprachlich handeln im Sachfach: Bilinguale Lernkontexte. In G. Bach, & J.-P. Timm (Eds.), *Englischunterricht. Grundlagen und Methoden einer handlungsorientierten Unterrichtspraxis* (4th Ed.) (pp. 172–198). Tübingen und Basel: Francke.

- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4th Ed.). Berlin: Springer Verlag.
- Bredenbröker, W. (2000). *Förderung der fremdsprachlichen Kompetenz durch bilingualen Unterricht: Empirische Untersuchungen*. Frankfurt am Main: Lang.
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany – Results from recent research in a contested field of education. *International CLIL Research Journal*, 1(4), 5–16.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL – Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe. In W. Delanoy, & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139–157). Heidelberg: Carl Winter.
- Evangelisch Stiftisches Gymnasium. (2015). Der bilinguale Zweig am Evangelisch Stiftischen Gymnasium Gütersloh. Retrieved from [http://www.esg-guetersloh.de/wp-content/uploads/2017/11/20660\\_flyer-bili.pdf](http://www.esg-guetersloh.de/wp-content/uploads/2017/11/20660_flyer-bili.pdf)
- Haagen-Schützenhöfer, C., Mathelitsch, L., & Hopf, M. (2011). Fremdsprachiger Physikunterricht: Fremdsprachlicher Mehrwert auf Kosten fachlicher Leistungen? *Zeitschrift für Didaktik der Naturwissenschaften*, 17(1), 223–260.
- Koch, A., & Bündler, W. (2008). Bilingualer Unterricht in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht*, 61(1), 4–11.
- Kondring, B., & Ewig, M. (2005). Aspekte der Leistungsmessung im bilingualen Biologieunterricht. *IDB - Berichte des Institutes für Didaktik der Biologie*, 14(1), 49–62.
- Kultusministerkonferenz. (2004). Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10). Retrieved from [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Biologie.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf)
- Lamsfuß-Schenk, S. (2008). *Fremdverstehen im bilingualen Geschichtsunterricht: eine Fallstudie*. Frankfurt am Main: Peter Lang Verlag.
- Landesinstitut für Schule und Weiterbildung in Soest. (1995). *Spotlight on History. Materialien für bilinguale Klassen*. Berlin: Cornelsen.
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1(1), 31–42.

- Müller-Schneck, E. (2006). *Bilingualer Geschichtsunterricht – Theorie, Praxis, Perspektiven*. Frankfurt am Main: Peter Lang Verlag.
- Osterhage, S. (2007). Sachfachkönnen (scientific literacy) bilingual und monolingual unterrichteter Biologieschüler: ein Kompetenzvergleich. In D. Caspari, W. Hallet, A. Wegner, & W. Zydati (Eds.), *Bilingualer Unterricht macht Schule: Beitrge aus der Praxisforschung* (pp. 41–50). Frankfurt: Peter Lang.
- Schmiedebach, M., & Wegner, C. (2017). Von der Handlungs- zur Bildungssprache – Beschulung neuzugewanderter Schler\*innen (Manuscript submitted for publication).
- Sprhase, U. (2012). *Biologie-Didaktik – Praxishandbuch fr die Sekundarstufe I und II* (5th Ed.). Berlin: Cornelsen Verlag.
- Wolff, D. (2011). Der bilinguale Sachfachunterricht (CLIL): Was dafr spricht, ihn als innovatives didaktisches Konzept zu bezeichnen. *Forum Sprache - Die Online-Zeitschrift fr Fremdsprachenforschung und Fremdsprachenunterricht*, 3(2), 75–83.
- Yassin, S. M., Tek, O. E., Alimon, H., Baharom, S., & Ying, L. Y. (2010). Teaching science through English: Engaging pupils cognitively. *International CLIL Research Journal*, 1(3), 46–59.
- Zydati, W. (2007). *DEUTSCH-ENGLISCHE ZGE IN BERLIN (DEZIBEL): eine Evaluation des bilingualen Sachfachunterrichts an Gymnasien*. Frankfurt am Main: Peter Lang Verlag.

## Appendix I

### Difficulty indices and discriminatory power of all items at $t_1$ (n = 72)

	Mean	Point score per item	Difficulty index (mean/point score)	Discriminatory power
Item 1	0.361	1	0.361	0.444
Item 2	0.708	1	0.708	0.167
Item 3	0.306	1	0.306	0.667
Item 4	0.417	1	0.417	0.389
Item 5	0.472	1	0.472	0.333
Item 6	0.389	1	0.389	0.667
Item 7	0.292	1	0.292	0.333
Item 8	0.042	4	0.010	0.000
Item 9	0.458	1	0.458	0.444
Item 10	0.597	1	0.597	0.222
Item 11	0.347	4	0.087	0.500
Item 12	0.569	1	0.569	0.667
Item 13	0.319	1	0.319	0.444
Item 14	0.806	4	0.201	0.667
Item 15	0.722	1	0.722	0.500
Item 16	0.806	1	0.806	0.389
Item 17	0.389	1	0.389	0.278
Item 18	0.403	1	0.403	0.333
Item 19	0.417	1	0.417	0.500
Item 20	0.222	1	0.222	0.222
Item 21	0.306	1	0.306	0.389
Item 22	0.569	1	0.569	0.333
Item 23	0.236	1	0.236	0.333
Item 24	0.056	4	0.014	-0.056
Item 25	0.167	3	0.056	-0.056
Item 26	0.847	3	0.282	0.444
Item 27	0.250	1	0.250	0.167
Item 28	0.472	4	0.118	0.833
Item 29	0.333	1	0.333	0.556
Item 30	0.333	3	0.111	0.500
Item 31	0.375	1	0.375	0.333
Item 32	0.208	1	0.208	0.111
Item 33	0.181	1	0.181	0.500
Item 34	0.014	4	0.003	0.056