

Foretelling or Fortuitous?

Analysing the English Admission Test to Bilingual Education at a Dutch Bilingual School



Utrecht University, Centre for Teaching and Learning, U-TEAch Course of August 2012
August, 2013

Authors: R.L. Burmanje, E.H. Deppe and F.G. Van Elzakker
Word Count: 6751

ABSTRACT

This research project questions the predictive values of an English admission test to a Dutch school with a bilingual programme (TTO). The main focus lays on the possible correlation between the results of the admission test and the results of the CLIL subjects, while also looking critically at the admission test itself. Relevant to all schools with bilingual programmes, this research shows teachers and other education professionals the significance or insignificance of admission tests. According to this project, the admission test does not have a predictive value for the term grades of the CLIL subjects.

INTRODUCTION

This practice based research questions the predictive values of an English admission test to bilingual education at a Dutch secondary school in the Netherlands. Each first year student enrolled in the bilingual programme has completed an admission procedure which consists of a paper written in Dutch, an interview in both Dutch and English and an English test. The focus of this research is on the English test. The school approached us with their problem concerning the admission test and its correlation with the results in CLIL (Content and Language Integrated Learning) subjects which are History, Biology and Geography. Some of the students appear to have problems in achieving good grades in the CLIL subjects and the case school is unable to identify a relation between the admission test and the grades obtained in the CLIL subjects. The school wanted us to find out whether the admission test has any predictive value for the first three terms of the school year. We planned to do this by analysing the admission test, mapping out the language skills used in the CLIL subjects (focus on reading, writing, listening, speaking and grammar) and by looking for a correlation between the grades obtained in the English admission test and the grades obtained for the CLIL subjects.

The research combines both qualitative and quantitative methods, as it looks at the language skills tested in the admission test and the CLIL subjects, but it also looks at term grades and admission test outcomes. These are processed through three different tests with SPSS Statistics (software). The qualitative methods are used to provide an explanation for the results drawn from the quantitative research.

Practical Relevance

In order to be admitted to bilingual education at the school, students have gone through an admission procedure set up by the school. The European Platform does not require a selection procedure for bilingual schools (O. van Wilgenburg, personal communication, May 29, 2013). This caused the school and perhaps other bilingual schools in the Netherlands to wonder how to best test future students. Specifically, the school wants to know whether they can use the test as a tool to predict students' future results in CLIL subjects, so they can offer support to the students at an earlier stage during the school year if this is indicated. As Anthonissen (2006) mentions, it is important that everyone involved in a test is aware of its consequences, which is currently not the case. With this research, we aim to provide a base for further research, because the use of admission procedures to bilingual education in the Netherlands has not been researched as of yet.

Theoretical Framework

This research deals with three main points; (1) the language skills that are tested in the English admission test, (2) the language skills that are used in the CLIL subjects and (3) the possible correlation between the grades obtained for the English admission test and the individual term grades obtained for the CLIL subjects. This research focuses on views from Anthonissen (2006) and Tarone (1999) concerning admission tests and the validity of testing specific skills, and on statements from Kwakernaak (2009), Bachman & Cohen (1999) on

English skill testing and its use outside of the classroom. Finally, Hajer (2000) and Dalton-Puffer (2008) argue the use of English skills within the (CLIL) classroom.

Anthonissen (2006) states that when using a test as part of an admission procedure, it needs to be clear what is tested exactly, and what the consequences are when someone performs below expectation (Anthonissen 2006). This means that a test needs to be transparent. In this case, the English admission test to bilingual education tests English grammar and English writing and reading skills. The test has no particular consequences for the students nor the teachers, whatever their performance, which raises questions about its reliability and validity. According to Anthonissen (2006), an admission test needs to have a clear connection to the test taker's language use in the (near) future (Anthonissen 2006). As Staatsen (2011) mentions, validity is an important aspect of a test and each test needs to measure what the designer wants to know. Also, it needs a goal that is clear to test-takers, makers and anyone else involved (Staatsen 2011). This leads us to consider which elements the admission test includes that might be used in the student's future high school career concerning the English language. In other words; which language skills are tested in the English admission test and what can be said regarding validity, reliability and transparency?

According to Tarone (1999), it is extremely difficult to test whether a learner lacking these skills equals that learner producing authentic language use, as the testing of just components or skills is not sufficient. It causes language variation, based solely on the specific test, and not on the proficiency of the test taker (Tarone 1999 & Shohamy 1999). Within the theories of language testing and language acquisition, there is a clear shift from testing and teaching components (like syntax or grammar) to testing and teaching language with a purposeful and communicative approach (Tarone 1999). These theories have not made it into practice yet, as Kwakernaak (2009) states that there often is a lack of correspondence within English tests and real life situations, as students may not always receive realistic language exercises, which can prepare them for their future life in society (Kwakernaak 2009). Bachman (1999) points out that it is not of importance how skilled someone is at producing correct grammar, but how well someone communicates at the needed level and context (Bachman 1999). These statements make us wonder whether this shift has also occurred at the school and consequently, what (language) skills are used in tests of the CLIL subjects that determine the term grade?

CLIL refers to a learning and teaching approach in which non-language subjects are taught through a foreign language. The focus of this approach lies in acquiring knowledge, skills and competences both in the subject and in the foreign language. Students enrolled in these CLIL subjects are required to learn a vast amount of content vocabulary as well as subject knowledge simultaneously. Hajer (2000) states that "it is feared that since the language used for learning is less perfectly known than the student's first language, it may lead to reduced subject competence either through imperfect understanding or through the fact that teachers anticipate this problem and simplify contents beforehand" (Hajer 2000). However, according to Dalton-Puffer (2008), research results on CLIL are generally positive, as most studies show that CLIL students have the same amount of content knowledge as their non-CLIL peers (Dalton-Puffer 2008).

As we are focused on admission test outcomes and term grades, and these two contrasting views only focus on the acquirement of knowledge and subject competence (Hajer 2000 and

Dalton-Puffer 2008), we are unable to use them to explain whether the use of a foreign language (English) within History, Geography or Biology has an effect on the student's term grades or not. This brings us back to the former two questions mentioned in this framework, and to the final question; is there a correlation between the grades obtained for the English admission test and the individual term grades obtained for the CLIL subjects?

Research Question

After consulting our contacts at the school and constructing the theoretical framework, the following research question was formulated: can the English admission test predict achievement in the CLIL subjects Biology, History and Geography? In order to find an explanation for the acquired outcomes and term grades, it was decided to explore the testing methods used in both the admission test and the tests for History, Geography and Biology. The outcomes may shed light on a possible correlation between the admission test and the CLIL subjects. To find out whether outcomes of the English admission test can predict achievement in the CLIL subjects, we decided to compare the outcome of the admission test to the term grades obtained in the CLIL subjects during the first three terms by using statistical analysis. This brought us to the following sub-questions: which (language) skills are tested in the English admission test and what are the test's validity, transparency, and reliability? What (language) skills are used in the CLIL subjects that determine the term grade? Is there a correlation between the grades obtained for the English admission test and the individual term grades obtained for the CLIL subjects?

Hypothesis

Because the ability to test someone's future language use is so important (Anthonissen 2006), we hope the admission test does exactly that. However, because it is so difficult to test authentic language use (Tarone 1999) and producing correct grammar does not show how well someone communicates (Bachman 1999), we arrived at the following hypothesis: if the English admission test only tests passive language skills, then it will not have a predictive value for History, Geography, and Biology in the first three terms.

Context

The school is a large Dutch secondary school in the Netherlands. It currently caters to about 1600 students, of which approximately 400 students follow bilingual education. The school started offering bilingual education in 2002 in *Atheneum* and *Gymnasium* and only admits students who scored at least 545 out of 550 points on their CITO test at the end of their primary education. The bilingual department usually allows for 75 students to enrol in the bilingual programme, but has admitted 79 new students in the school year 2012-2013. Of these 79 students, 7 are native speakers of English and 72 are non-native speakers of English. Students may have different starting levels of English but also varying intellectual abilities. Some students may be quite adapt at certain subjects such as Geography or History, but less so in Biology. Also, students who are good at English may not be good at CLIL subjects necessarily and vice versa. The admission procedure takes place approximately six months before the start of the school year. The procedure consists of a paper written in Dutch, an interview in both Dutch and English and an English test. The English test is part of the

procedure and is created by the English department. The test aims to test English proficiency and consists of six different components: (A) Reading test, (B) Unscramble, (C) Put words in the right order, (D) Circle the correct answer, (E) Reading and understanding and (F) Writing assignment. The entire admission procedure takes about two and a half hours; there is no set time for the English test. A maximum score of 42 points can be obtained. This test does not test speaking or listening skills.

The points obtained for the admission test are compared to all the grades the student has obtained in the CLIL subjects throughout the first three terms. Term 1 ran from September 6th until November 21st, Term 2 from November 22nd until January 30th and Term 3 from January 31st until April 19th. The average grade obtained in Term 1 also contributes to the grades obtained in Term 2 and the average grade from Term 2 contributes to the grades obtained in Term 3.

METHOD

Sub-question 1: *which (language) skills are tested in the English admission test and what are the test's validity, transparency, and reliability?*

Variables

Language Skills

A language skill is a component of a language that can be tested in order to assess someone's proficiency in that language. It is likely that the language skills reading, writing, listening, speaking and grammar are tested in this admission test, as these are the most common language skills used within an artificial language acquisition environment, in this case language acquisition within the classroom.

Validity

A test becomes valid, i.e. actually tests what it aims to test, when its goals are clear and they are attainable or measurable by means of the test (Staatsen 2011).

Transparency

A transparent test is a test in which outcomes and consequences are clear to everyone involved: not only a test maker and taker but also the people involved in a test situation, such as future teachers, assessors of the test or parents of the test takers (Anthonissen 2006).

Reliability

Reliability of a test has to do with clarity and the absence of ambiguity within assessment schemes, assessors and what the consequences of outcomes are (Staatsen 2011).

Data Source

The test (appendix A) and the information regarding the test were provided by a staff member at the school, who is also the maker and assessor of the test, i.e. the test items and the assessment scheme.

Procedure

We analysed the admission test by looking at theoretical approaches to language testing and exercises by Staatsen (2011) and Kwakernaak (2009). They had not come up with an instrument for test analysis, but by comparison their examples acted as a tool in our analysis. Through this, we categorised the six different components of the test into language skills tested. We also used the examples of Staatsen (2011) and Kwakernaak (2009) to find out which performance and comprehension levels come into play with the selection of exercises, to see how these relate to each other and to the goals of the test.

Methods of Analysis

For the validity, transparency and reliability we based our analysis on Staatsen (2011). This was mainly focused on the perceived goals of the test, as well as what is actually measured within each item. We also analysed the admission test to determine if it has a clear and unambiguous assessment scheme, multiple assessors, and whether outcomes are discussed, as per Staatsen (2011).

Sub-question 2: which language skills are used in the CLIL subjects that determine the term grade?

Variables

Language Skills

In our analysis of the admission test, we focus on the language skills (see sub-question 1) reading, writing, listening, speaking and grammar. For comparison, we created a questionnaire for the CLIL teachers in which we focus on these language skills. These skills may be tested in the classroom, as they are the most common language skills used within an artificial language acquisition environment.

Respondents

Teachers of the CLIL Subjects

The teachers of the CLIL subjects who teach the first forms (1A, 1B and 1C) at the Dutch bilingual school were requested to fill out a questionnaire. There are four teachers in total; one Geography teacher, one Biology teacher and two History teachers. These teachers, aged between 28 and 41, all have several years of experience in teaching in general and have been teaching in bilingual education between three and nine years. The questionnaire contains the answers of four CLIL teachers, which makes the response anecdotal. This factor reduces the reliability of this research.

Instruments

Questionnaire

We created a questionnaire for the teachers of the CLIL subjects. It consisted of 24 questions; 13 open and 11 multiple choice questions. The questions focused on the teachers' knowledge and use of CLIL, the role of English language skills (reading, writing, listening, speaking, grammar, vocabulary and fluency in particular) and the admission procedure (appendix B). These questions aimed to find out whether there is a similarity between the English skills that

are most used and tested in the CLIL subjects and the skills that are tested in the English admission test. The responses would help to explain the possible link between the term grades and the outcomes of the admission test (appendix B and C).

Procedure

Pilot Questionnaire

Two weeks before sending the online questionnaire to the teachers, we sent out an online pilot questionnaire to our peers, as they are familiar with CLIL philosophy. Our peers provided us with feedback which we could use to take out the bugs of the instrument. This would provide us with a complete and valid questionnaire, which the actual respondents can complete without experiencing difficulties (Bell 2010). Six peers filled out the questionnaire and provided us with online feedback. Based on this, we rephrased ambiguous questions and left out questions that turned out to be redundant. It took our peers approximately 20 minutes to complete the questionnaire.

Questionnaire

We sent out the questionnaire (www.thesistools.com) online to four teachers who teach CLIL subjects in the first form. The teachers were able to fill out the questionnaire in their own time and we gave them a week to complete it. Two out of the four completed the questionnaire within a week, two others needed a reminder. After two weeks, all four questionnaires were completed.

Method of Analysis

We analysed the questionnaire by gathering the answers to the questions and by categorising them. The categories focused on the following; questions on profession, use of CLIL, the role of English language skills and how the admission procedure is used. We compared the relevant outcomes (focus on use of CLIL and the role of English language skills) and based on these and possible recurring answers we formulated a conclusion for our sub-question.

***Sub-question 3:** is there a correlation between the results of the admission test and the individual term grades of CLIL subjects?*

Variables

Admission Test Outcomes

The first variable is the outcome of the admission test to the bilingual programme: it is a test score between 0 and 42 points awarded to the learners responses. The outcomes consist of the results of the 79 learners who have participated in the English test in the admission procedure in 2012.

Term Grades

The second variable, the term grades, are made up of the up to 1 decimal average of all the term assessments, administered in the CLIL subjects throughout the first three terms of the school year. The average grade obtained in Term 1 also contributes to the grades obtained in

Term 2 and the average grade from Term 2 contributes to the grades obtained in Term 3. It is unknown how many tests make up a term grade.

Respondents

Bilingual Programme First Year Students

The final three term grades for the CLIL subjects from the 79 first year students in the bilingual programme in 2012 were sent to us. We also obtained the admission test outcomes of these students. The outcomes and grades were made anonymous by categorizing students per student number.

Instruments

Three different statistical instruments were used to analyse the data, providing us with three independent outcomes. By triangulating these outcomes, we aimed to reach more reliable and valid conclusions.

Pearson's Correlation Coefficient, Independent T-Test, Regression Analysis

In order to analyse the term grades and admission test outcomes, we used IBM SPSS Statistics.

The first instrument was the Pearson's Correlation Coefficient to examine the presence of a correlation between the term grades and admission test outcomes. The Pearson's Correlation Coefficient, also known as Standard Correlation, which "examines whether two variables [...] correlate with each other and thus have consistency" (De Vocht 2011), was used to compare the outcomes of all 79 admission tests with all the grades from the CLIL subjects obtained in the first three terms.

Secondly, the Independent T-Test was used to compare the (variation in) test outcomes and term grades of two sample groups at the two extremes of the achievement ranges. In other words, to test if the best/least admission outcomes in fact match the best/least term grades. We created two sample groups from the admission test outcomes; a group with the highest (8) outcomes and a group with the lowest (9) outcomes.

The third instrument, the regression analysis, compared the (variation in) in admission test outcomes with the (variation in) term grades once again, but now for the whole student group and this time per CLIL subject and per term. This last test ensured no results would remain undetected.

Method of Analysis

Standard Correlation

Each combination of admission test outcomes with a CLIL subject term grade was plotted against each other and checked for outliers, as these may influence the outcome of the data (see example in appendix D, Fig.1.0). Major outliers will be omitted.

Independent T-Test

With the Independent T-Test, we tested the differences between two averages; the lowest 9 admission test outcomes (group 1, between 23 and 30 points) and the highest 8 admission test outcomes (group 2, between 42 and 40 points). These two groups are both relatively

homogenous in the amount of points and the group size differs minimally. In this way, we prevented an above average outcome in one group to increase the average of that entire group. We used Levene's Test of Equality of Variances to check if the variances of the two groups were equal (homogeneity of variances), which is the case if the significance level (p -value) is greater than 0.05 (Laerd Statistics 2013). This p -value checks whether or not results are coincidental. We chose the Independent Samples T-Test, which produces a table (appendix D, Table 1) showing the significance of the admission test outcomes compared to the term grades for each group in the (*Sig. (2-tailed)*) column.

Regression Analysis

An analysis of the regression is done on the correlation of the admission test outcomes (independent variable) and the grades for a CLIL subject from one term (dependent variable). According to De Vocht (2011), the simple regression analysis says something about the relationship between the independent and dependent variables, but not everything, as chances are that this relationship is misinterpreted. To prevent this, it is important to ensure that the data that is used is continuous or ordinal (De Vocht 2011) as is the case in this research: both the admission test outcomes and the term grades are ordinal.

The regression analysis offers three values per tested relationship (see example in appendix D, Fig. 3.0). The Model Summary table, which shows the R value that represents a simple correlation, and the R² value, which "indicates how much of the dependent variable (in this case CLIL subject term grade) can be explained [or predicted] by the independent variable (in this case admission test outcomes)" (Laerd Statistics 2013). The ANOVA table indicates whether the regression model predicts the outcome variable significantly well. This can be seen in the regression row and Sig. column. If the value is less than 0.05, it means that the applied model can "statistically significantly predict the outcome variable" (Laerd Statistics 2013). Finally, the Coefficients table "gives us the information we need to predict [a CLIL subject term grade] from [the admission test]" (Laerd Statistics 2013). If the values in the Sig. column and in the admission test ("intake") rows are $p < 0.05$, we can say that the admission test contributes significantly to the CLIL subject term grade.

If the Standard Correlation shows correlation, the Independent T-Test and the regression analysis are significant, this shows that the admission test is predictive for the grades of CLIL subjects in the first year of bilingual education.

RESULTS

Sub-question 1: *which (language) skills are tested in the English admission test and what are the test's validity, transparency, and reliability?*

Language Skills

The admission test (appendix A) consists of six assignments, which focus on reading (assignments A and E), writing (F), and grammar (B, C, D). Both A and E (reading comprehension) and B and C (unscramble sentences) are similar. In fact, there are only four kinds of assignments within this admission test: two reading comprehension texts, two

sentence-unscramble exercises, one set of multiple choice questions ('circle the correct answer'), and a simple writing assignment with a focus on spelling and grammar. The first three kinds all require passive language skills, while the latter is the only one that focuses on active skills.

Validity, Transparency and Reliability

From questions we have asked the contacts at the school, we have concluded that the validity of the test is impossible to determine, as the designer was unable to say what they want to know and therefore whether that is measured. This is based on the test itself and the fact that the test results are not used in any way. A student is not rejected or given any scaffolding when they score low on this test and the results are not discussed with other staff. The test is not transparent, as its goals and consequences are unclear to takers, makers, and everyone involved. Since no rubric or assessment scheme was used to grade the test, and it was graded by one school official this affects the reliability of the data source.

Sub-question 2: *which language skills are used in the CLIL subjects that determine the term grade?*

The Use of CLIL

All teachers of the CLIL subjects state that they know what CLIL entails and they claim to actively apply CLIL in their classrooms. In Biology, cross-curricular projects are organised in which teachers for English and for Biology are present during the final presentations. Students receive a mark for content as well as a mark for their use of English. The History teacher claims that "less effort is put in designing CLIL activities than before, due to lack of time" (appendix C). The other History teacher mentions that CLIL is practiced through many different learning activities, such as use of visuals (simplifying the content), splitting up the more difficult texts into chunks and the use of non-linguistic tasks such as storyboards and posters.

The teachers' experience with CLIL in terms of English acquisition from (previous) experience is positive; the Biology teacher mentions that it is the "best way of learning [...] students experience that subject teachers are language teachers as well" (appendix C). In addition, all teachers show a positive attitude towards CLIL as a concept, as three out of four grade it as 'good' and one thinks it is 'excellent' (appendix C).

The Role of English Language Skills

According to the teachers of History and Geography; reading, listening and speaking skills play the most important role in regarding their subject. The Biology teacher mentions that all skills are equally important. When it comes to marking language mistakes, the teachers have a different approach. The Geography teacher and one of the History teachers do not mark language mistakes, while the other History teacher sometimes marks language mistakes. The Biology teacher does mark language mistakes and emphasises this by stating "I teach history and not English" (see appendix C). Three out of four do not assess students on their use of English in tests, whereas one History teacher sometimes assesses students claiming that "it is necessary for their language skills to develop, which in turn enables the students to have a

better understanding of my subject” (appendix C). Three out of the four teachers declares that they are not aware of the language skills that are tested during the admission procedure.

Sub-question 3: *is there a correlation between the grades obtained for the English admission test and the individual term grades obtained for the CLIL subjects?*

Standard Correlation Test

In Fig 1.0 the correlation coefficients are shown between the admission test and the three term grades for the three CLIL subjects. For definite correlations the coefficient is marked with a double asterisk, medium correlations with a single asterisk. Unmarked coefficients indicate absence of a correlation. The first term grades show only for History a definite correlation (.347), no correlation for the Biology and Geography. The second and third term grades are moderately correlated for all three subject to the admission outcomes. However, in the third term the History term grades show a definite correlation (.303) once again. The graph also shows that for each term, the correlation coefficients are lower for Biology and Geography than for History.

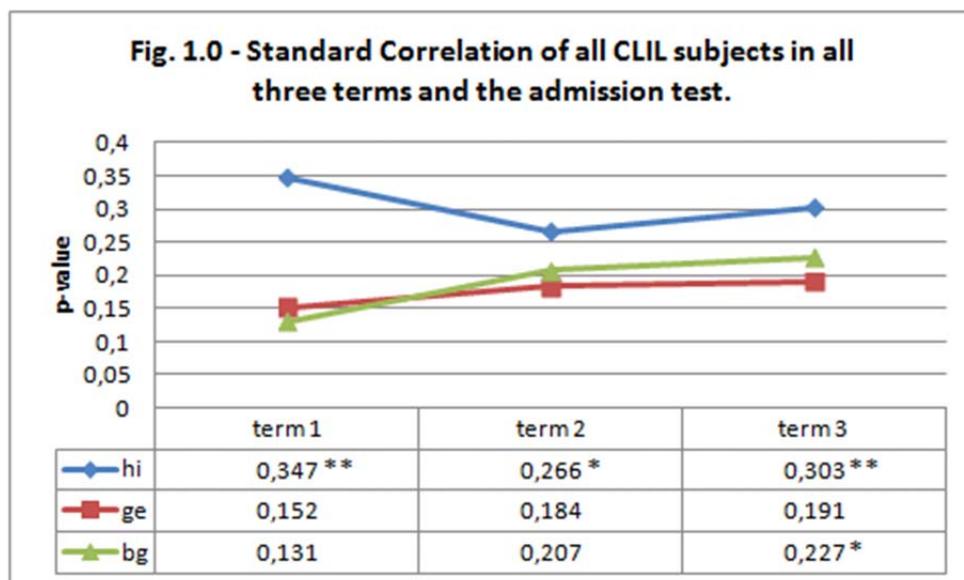


Fig 1.0. p-values (correlation) found during the Standard Correlation analysis between admission outcomes and each CLIL subject in all three terms.

Independent T-Test

In this test, the group of participants with the lowest admission scores is plotted against the group of participants with the highest admission scores, to see if they achieve significantly less for their term grades in the three CLIL subjects. For all three subjects, both sample groups (the top 8 admissions and the 9 least admissions) show more or less the same patterns of significance, when the sample averages of the admission is tested with the average term grades of that sample group. The graphs of p-values versus term grades for all three subjects can be found in graphs 2,0, 2.1, and 2.2 showing all p-values to be greater than 0.1.

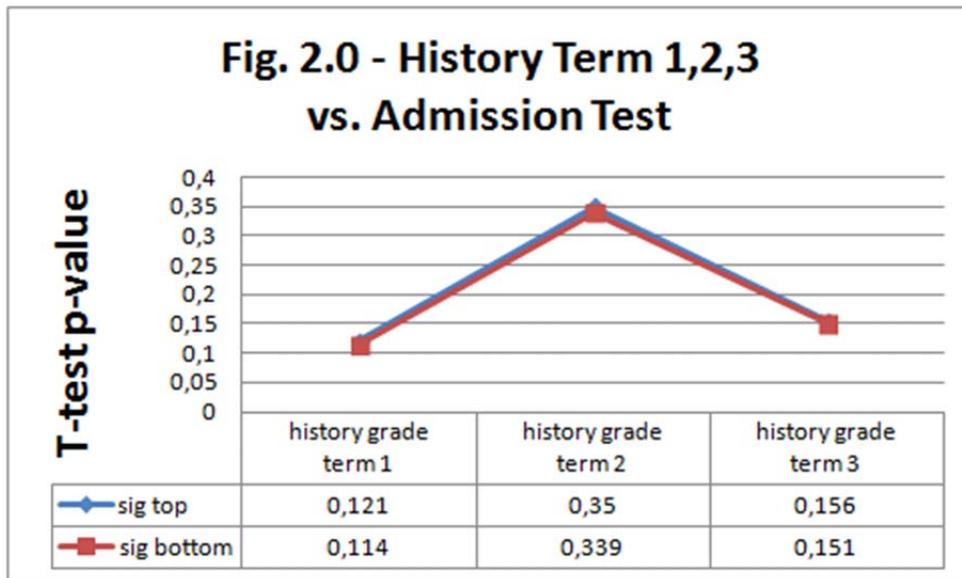


Fig 2.0. p-values (significance) found during T-test analysis between admission outcomes and the three term grades of the CLIL subject History.

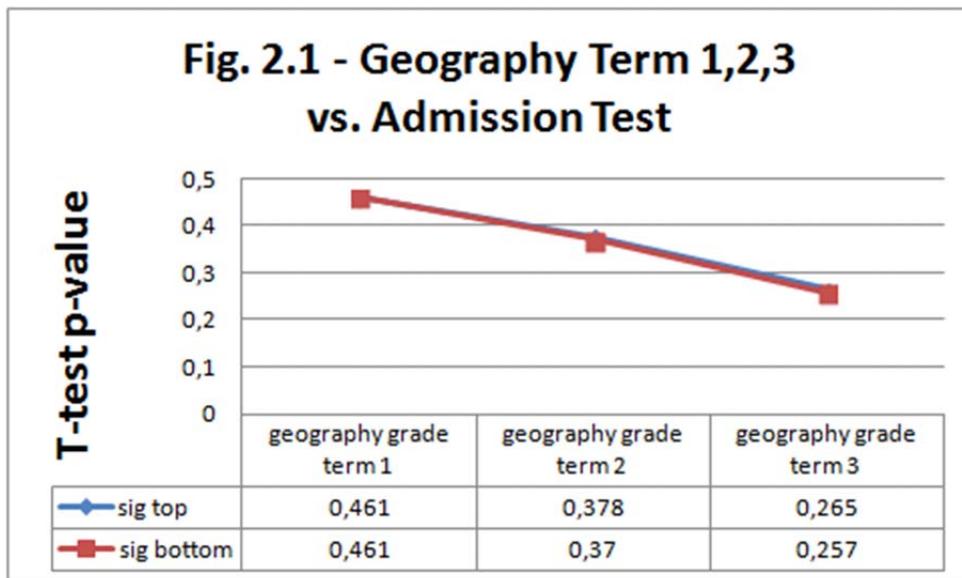


Fig 2.1. p-values (significance) found during T-test analysis between admission outcomes and the three term grades of the CLIL subject Geography.

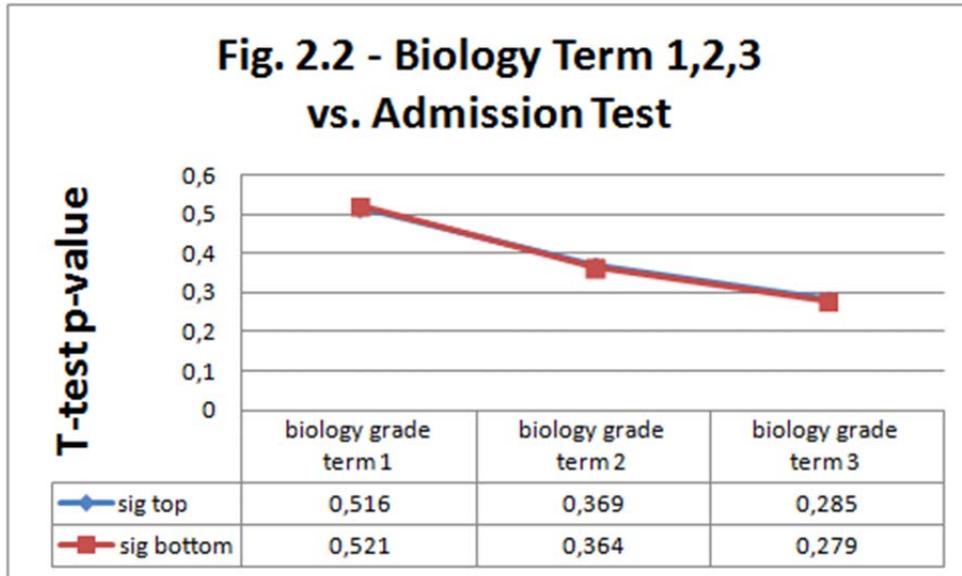


Fig 2.2.. p-values (significance) found during T-test analysis between admission outcomes and the three term grades of the CLIL subject Biology.

Regression Analysis

Table 1 below shows the p-values obtained in the regression analysis for the three subjects and three terms. As the R^2 -values in this correlation are all (except in one case) below the 0.1 standard defined in the methodology, there the CLIL subject term grades cannot be explained by the admission test outcomes.

Subject	R^2 -values Term 1	R^2 -values Term 2	R^2 -values Term 3
History	0.120	0.071	0.092
Geography	0.023	0.034	0.036
Biology	0.017	0.043	0.052

Table 1: p-values for the regression analysis of each CLIL subject and term.

The graph in Fig 3.0, shows p-values (significance) of the R^2 found during regression analysis between admission outcomes and the three term grades of three CLIL subject (n=79). The p-values for all terms of History, all satisfy $p < 0.05$, which means that the term grades of this subject can possibly be predicted by the admission test. Geography and Biology do not show any significance apart from Biology's last term grade, as seen in Figure 3.0.

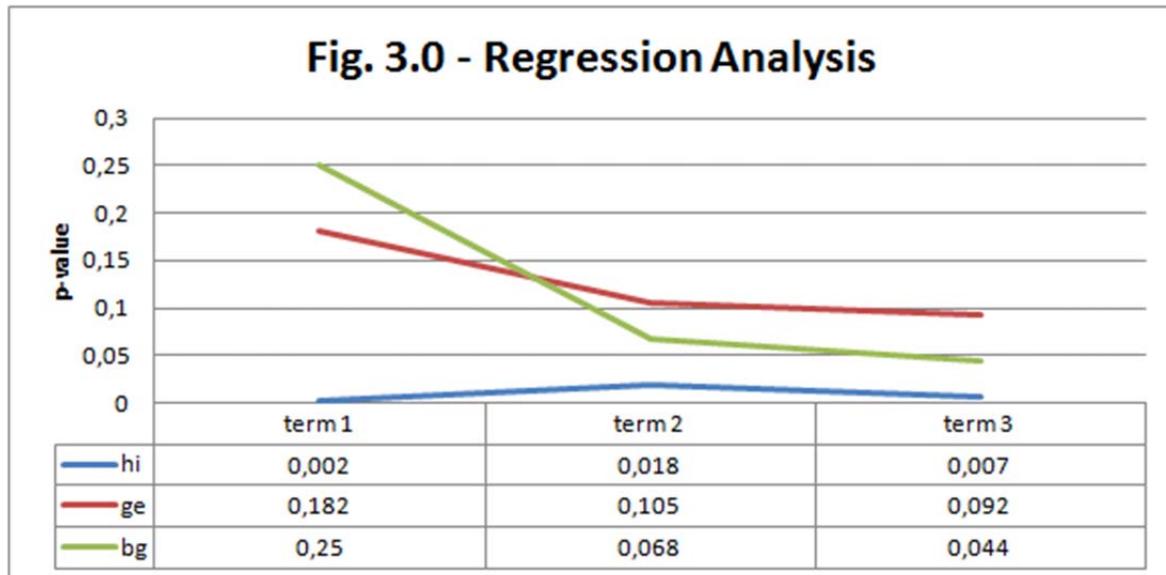


Fig 3.0. p-values (significance) of the R^2 found during regression analysis between admission outcomes and the three term grades of three CLIL subject (n=79).

CONCLUSION

Sub-question 1: which (language) skills are tested in the English admission test and what are the test's validity, transparency, and reliability?

In the English admission test, only some language skills are tested in various ways: simple reading comprehension (true/false statements), sentence unscrambling, a multiple choice grammar test and a writing assignment with a focus on correct spelling and grammar. The focus of this admission test is on form instead of on communication. It does not include a speaking or listening component and thus focuses mainly on passive skills.

Because there is no clear goal to this admission test, and no consequences attached to set outcomes, this test may be considered invalid. The test may not measure what the designer wants to know (i.e. how a student will perform in CLIL subjects). On the other hand, the assessment of the test may be considered reliable, as it is done by one person. However, these corrected tests and their outcomes are never discussed with or communicated to anyone, meaning there is a lack of transparency.

Sub-question 2: which (language) skills are used in the CLIL subjects that determine the term grade?

Based on the response of four subject teachers at this school we can say that in Biology History and Geography active communicative languages skills i.e. reading, listening and speaking are considered vital. In Biology writing skills are also considered equally important. Passive language skills such as grammar, vocabulary and fluency do not appear to play an important role in the CLIL subjects.

The teachers do not assess the students on their use of English in tests. This means that the (term) grades the students obtain are not influenced by the language mistakes they might make. Nevertheless, weak reading and listening skills may cause the students to

misunderstand content. These skills most likely influence the term grade for these CLIL subjects.

Sub-question 3: *is there a correlation between the grades obtained for the English admission test and the individual term grades obtained for the CLIL subjects?*

Combining the results for each statistical test, we can conclude the following:

- 1) Only the History Term 1 grade shows a definite standard correlation with the admission test outcomes.
- 2) The top 8 and lowest 9 in the admission do not correspond significantly to the top 8 – lowest 9 in the three CLIL subjects, not even for History Term 1.
- 3) The ANOVA from the regression analysis indicates that the admission test outcomes cannot significantly predict any of the term grades in the three CLIL subjects.

Therefore, we can state that there is not a significant correlation between the grades obtained for the admission test and the individual grade terms obtained for Biology and Geography. There is an indication that term grades of History do correlate with the admission test, however, this is not consistent per test and term, which leads us to dismiss the possibility that the outcomes of the admission test can predict the term grades of History. In sum; the current admission test outcome does not predict a student's future results in CLIL subjects.

DISCUSSION

This research has shown that the grades from the first year students of the school year 2012/2013 could not be predicted by the admission test. Based on the data analysis and theoretical framework, we are not able to disprove our hypothesis.

Discussion of the Research Quality

Admission Test

The analysis of the admission test has been done based on theory from Staatsen (2011) and Kwakernaak (2009). By primarily focusing on these two authors, our theory may be quite limited to the Dutch point of view. As Staatsen and Kwakernaak do not mention an instrument for test analysis, we used their descriptions of test items to form our own test analysis through comparison. Though we did focus on the concepts of validity, transparency and reliability when analysing the test, our approach is not necessarily an accurate way of analysing a test as it was not systematic enough. We were able to compare examples found in theory to the admission test but did not have an unambiguous way to measure the validity of the test. We could have interviewed our school contacts more thoroughly about the goals and communications concerning the test. However, the analysis of the admission test itself is unambiguous and almost exclusively tests passive skills.

Questionnaire

The small amount of respondents makes it hard to generalize the outcomes of the questionnaire. Also, we lack information due to the fact that some teachers did not answer all

the questions or did not explain their answers. We think that this may have happened because some teachers did not know how to answer certain questions, they did not know how to use the online questionnaire or because they were not interested in the research.

Furthermore, as most of the respondents knew what the research was about and were aware of the fact that the results would be published once finished, there is a chance that the respondents were biased. They might want to please their employers or the researchers, which can influence the reliability of the answers provided. However, the outcome of the questionnaire did not influence the researched outcome, which is why the questionnaire did not bias the research as a whole.

All in all, although at first we thought an online questionnaire would be a good tool to use for our research, it proved to be not as useful while we were analysing the results.

Data Analysis

The analysis of the data can be considered reliable, as it has used three different well-known statistical tests to process the data. Similar data, if processed under the same circumstances will yield similar results, making the analysis generalizable to further studies. However, the analysis is purely based on the outcomes and grades from one school year. By analysing data from consecutive years, the analysis could be more valid and generalizable as well. By using the grades and outcomes from different first year students in subsequent years, the possibility of chance can be diminished. This would also give insight in to whether the medium correlation and significance found between the History term grades and the admission test outcomes is coincidental or if the variation in correlation is indeed a pattern.

Besides comparing the admission test outcomes to the CLIL subject term grades, we also compared the admission test to the English term grades. If a high correlation is found, this would mean that the validity of our approach (our choice in tests) would be justified, and that the admission test is indeed capable of predicting future grades, though not for all subjects. We performed the same three statistical tests used for this research with the admission test outcomes and the English term grades, and found a strong, significant correlation. This would be expected as the admission test contains mostly passive language skills which are probably also tested in the English term grade. Above all, this suggests that the chosen statistical approaches are capable of identifying predictors and the conclusions we arrived at concerning the CLIL subjects are justified. We found a correlation were we had expected to find one, which proves that our research method works.

Suggestions for Further Research

During the research we used the English term grade to verify the method of data analysis. The admission test outcomes and the English term grades showed high correlation. This suggests that the outcome of the admission test does have a predictive value for English. With this in mind, a different approach to this research might be to compare the term grades obtained in English to the term grades obtained in the CLIL subjects and compare both to the admission test outcomes. This may provide some clarity on the role that the English language plays in the CLIL subjects and term grades, but also how the admission procedure might be redeveloped so it can play a predictive role for the CLIL subjects.

Suggestions for the Practice

The admission test does not have a predictive value for the term grades of the CLIL subjects, nor does the outcome of the admission test have any consequences for the students or is it discussed or used by the CLIL and English teachers.

Hence, it might be advisable to change the admission test and procedure. The easier solution is to continue selecting students by CITO score, and still have the motivational talk, but leave out the admission test. However, it would be advisable to change the test in such a way that it has a diagnostic effect and discuss the outcomes with teachers of weaker performing students. The idea of being able to identify possible learning problems within the group of starting students to be able to support them at an earlier stage seems laudable. Thus it would be advisable to use the admission test in this way. Another suggestion could be to change the current admission procedure completely and create a new procedure based on the goals of the school: a possible predictive value, establishing a student's motivation, and possibly attaching consequences to student selection. A way to create a more valid and predictive admission procedure could be to do this with a team of CLIL and English teachers. This way, it will be more related to CLIL subjects and English, unlike the current admission test, which only has a strong correlation with passive English language skills.

We believe that our research is a valuable addition to contemporary research, connecting literature on CLIL and language tests. This research can therefore be used by other bilingual schools who want to discover how this specific admission test relates to future achievements in CLIL subjects. This research can be used as general advice towards admission procedures, but is not applicable to other admission tests and selection procedures into bilingual education. The admission test content and the use of English in CLIL subjects at other schools may differ significantly, as well as the level of students' capabilities and the composition of native and non-natives speakers within one group.

Reflection

This practice based research has been a very interesting experience. We worked with IBM SPSS Statistics and seen what definite results it can offer. This is a very rare occurrence as we are literary educated students. The process of setting up a research plan, creating and analysing a questionnaire, discussing research questions with the school, conducting the research with three different statistical tests and finally writing this article have all been a very educational and enriching experience. We feel that these tools will be useful for further possible researches we might want to do in our future careers. We also believe that this research will help us in the future, when reading scientifically based research on literary or education. We are proud to present our final product, which has been a large step outside of our comfort zone. However, we would recommend future research groups to mix up alpha, beta and gamma students, as having experts in different fields of scientific research has proven to be essential to carry out the research.

Endnote

R.L. Burmanje, E.H. Deppe and F.G. Van Elzakker are student teachers for the subject of English at the Centre for Teaching and Learning of Utrecht University. To qualify as teachers of upper secondary schools they conducted a research concerning the admission procedure to bilingual education at a Dutch bilingual school. The results of this Practice Based Research and the subsequent recommendations made provide the grounds for the above article.

REFERENCES

- Anthonissen, C. (2006). On Determining what Counts while Counting: Aspects of Language Testing where Diversity is the Standard. *Per Linguam*, 22, 39-54. doi: 10.5785/22-1-60
- Bachman, L. F., & Cohen, A. D. (1999). Language Testing - SLA Interfaces: an Update. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between Second Language Acquisition and Language Testing Research* (pp. 1-31). Cambridge: Cambridge University Press.
- Bell, J. (2010). *Doing Your Research Project: A Guide for First-Time Researchers in Education, Health and Social Science*. Maidenhead: Open University Press.
- Boeije, H. R., 'T Hart, H., & Hox, J. (2005). Enquete. In *Onderzoeksmethoden* (pp. 228-261). Den Haag: Boom Onderwijs.
- Dalton-Puffer, C. (2008). Outcomes and Processes in Content and Language Integrated Learning (CLIL): Current Research from Europe. In W. Delanoy & L. Volkman (Eds.), *Future Perspectives for English Language Teaching* (pp. 139-157). Heidelberg: Carl Winter.
- Ellis, R. (1994). Individual Learner Differences. *The Study of Second Language Acquisition* (pp. 471-528). Oxford: Oxford University Press.
- Hajer, M. (2000). Creating a Language-Promoting Classroom: Content-Area Teachers at Work. In J. K. Hall & L. Stoops Verplaetse (Eds.), *Second and Foreign Language Learning through Classroom Interaction* (pp. 265-85). London: Lawrence Erlbaum Associates.

- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press Inc.
- Kwakernaak, E. (2009). *Didactiek van het Moderne Vreemdetalenonderwijs*. Bussum: Uitgeverij Coutinho.
- Laerd Statistics. (2013). *Linear Regression Analysis using SPSS*. Retrieved June 10, 2013, <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>
- Rixtel, van J. (2009). Questionnaire on ThesisTools. Retrieved May 17, 2013, <http://www.thesistools.com/web/?id=346719>
- Shohamy, E. (1999). How Can Language Testing and SLA Benefit from Each Other? The Case of Discourse. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between Second Language Acquisition and Language Testing Research* (pp. 156-176). Cambridge: Cambridge University Press.
- Staatsen, F. (2011). *Moderne Vreemde Talen in de Onderbouw*. Bussum: Uitgeverij Coutinho.
- Tarone, E. (1999). Research on Interlanguage Variation: Implications for Language Testing. In L.F. Bachman & A. D. Cohen (Eds.), *Interfaces between Second Language Acquisition and Language Testing Research* (pp. 71-89). Cambridge: Cambridge University Press.
- Vocht, de A. (2011). *Basishandboek SPSS 19*. Utrecht: Bijleveld Press.

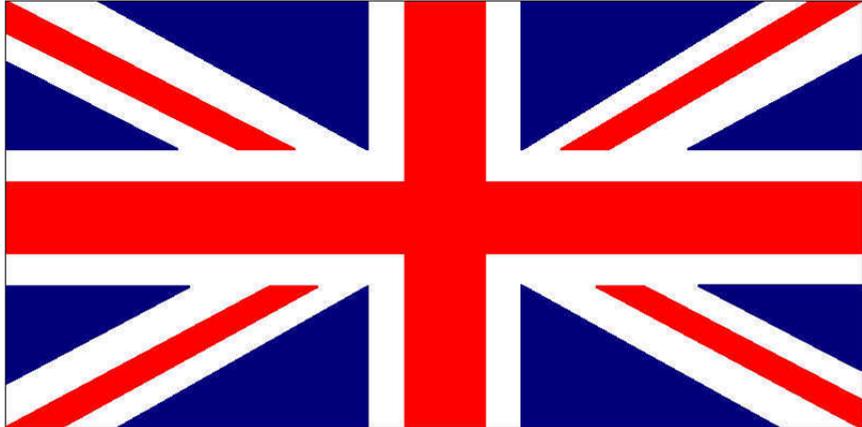
APPENDIX A – ADMISSION TEST

ENGLISH LANGUAGE TESTING

Name:.....

Test

Result:.....



Good luck!!!

A. Reading Test

MR JOHNSON

Mr Johnson looks at his watch. It is half past seven. He gets out of bed quickly, takes a shower and gets dressed. He is late as usual, so he does not have time for breakfast. He runs all the way to the station and he arrives just in time for the train.

Mr Johnson never eats anything in the morning. He always says to his friends at the office: 'It's nice to have breakfast in the morning, but it is nicer to lie in bed!'

A. YES OR NO:

1. Mr Johnson does not eat breakfast.

2. Mr Johnson goes to work at half past seven.

3. Mr Johnson is usually late getting up.

4. Mr Johnson takes the car to work.

5. Mr Johnson likes to get up early and eat breakfast.

B. Unscramble (zet in de goede volgorde):

1. has my sister at breakfast 7.30

.....
.....

2. a coffee and she a sandwich has

.....
.....

3. goes to by she work bike

.....
.....

Score:

	8
--	----------

C. Put the words in the right order

1. is / Brian's / bike / green

2. these? / are / Whose / shoes
.....

3. Zoe's / It's / bag

4. that? / is / jacket / Whose

5. dress / Emma's / pink / is

Score:

	5
--	---

D. Circle the Correct Answer

1. Is it an elephant?

- A. Yes, it's. B. Yes, it is. C. Yes, she is.

2. I short.

- A. amn't B. aren't C. 'm not.

3. They friends.

- A. is B. 're C. isn't

4. eight?

- A. She is B. She's C. Is she

5. Are you here?

- A. Yes, we are. B. Yes, they are. C. Yes, you are.

6. isn't funny!

A. We

B. You

C. He

7. You are

A. good singer

B. good singers

C. a good singers

8. Dave tall.

A. 's

B. aren't

C. he's

9. students?

A. Is it

B. They are

C. Are they

10. It isn't

A. bird

B. a bird

C. birds

Score:

	10
--	-----------

E. Reading and Understanding



Hard work and
no money

Every year thousands of young people in Britain finish school and then take a year off before they start university. Some young people go to other countries and work as volunteers. Volunteers give their time to help people – for example, they work in schools or hospitals, or they help protecting animals

Are the sentences True (T) of False (F)?

1. Pauline comes from Belize.
2. Pauline wants to save the coral reefs.
3. Pauline is studying the fish in the sea near Belize.
4. Pauline is unhappy, because she doesn't get any money.
5. Pauline wants to go home when she finishes her work.

Pauline Jones, 18, lives in London. Next year she wants to go to university to study Spanish. But now she is living in Belize, near Mexico. Pauline says, "I'm working with other people here to protect and save the coral reefs in the sea. The reefs are beautiful, but the sea water is very polluted*, the coral dies. I'm helping to do research* on the coral and the fish that live around the reefs. All over the world, coral reefs are dying. We need to do something about the problem before it is too late.

I'm staying with a family here and I help with the cooking and the cleaning. I don't get any money, but that is okay. I love my work here, and I am learning a lot about the people of Belize – and myself! When I finish my work, I want to stay here for another three months. I want to travel around Belize, Mexico and Guatemala.

* polluted = vervuild

* research = onderzoek

Score:

	10
--	----

F. Writing Assignment

NAME:

WHAT DO YOU DO IN THE MORNING? Describe your daily routine **in English**.

(schrijf minstens 5 hele **Engelse** zinnnetjes!!; meer mag ook))

.....

.....

.....

.....

.....

.....

.....

Score:

	10
--	----



APPENDIX B - QUESTIONNAIRE

CLIL at Hermann Wesselink College

1.

What is your age?*

2.

Which subject do you teach?

History
 Biology

Geography

Other

3.

Which class(es) do you teach?

1A
 1B
 1C

4.

How many years have you been a teacher?*

5.

How many years have you been teaching in bilingual education (at HWC and/or other schools)?*



The following questions are about the use of Content Language Integrated Learning within your subject. CLIL is an approach for learning (non-linguistic) content through an additional language (foreign or second), thus teaching both the subject and the language. CLIL is fundamentally based on methodological principles established by research on "language immersion".

6.

Have you been teaching according to CLIL philosophy? If so, since when?*

7.

What is your definition of CLIL and how do you apply CLIL in your classroom?*

8.

How do you experience CLIL in terms of English acquisition, in general (i.e. at other schools, previous jobs)?*

- Excellent
- Good
- Neutral
- Bad
- Terrible
- Not applicable

9.

Please explain why:*

10.

What is your opinion on CLIL?*

- Excellent
- Good
- Neutral
- Bad
- Terrible

11.

Please explain why:*

These questions focus on the role of English language skills and assessment in CLIL subject tests.

12.

Which English skills are most important during your lessons?

(Please number the skills in the order of importance. 1= most important, 5 = least important)

	1-5
Reading	<input type="text" value=""/> ▾
Writing	<input type="text" value=""/> ▾
Listening	<input type="text" value=""/> ▾
Speaking	<input type="text" value=""/> ▾
Other	<input type="text" value=""/> ▾

13.

Please explain why:*

14.

Do you mark language mistakes in learners' work?*

- Yes
- Sometimes
- No

15.

Please explain why (not):*

16.

Are learners assessed on their English in your tests?*

- Yes
- Sometimes
- No

17.

If yes, what are they assessed on?

- Reading comprehension
- Listening comprehension
- Speaking
- Writing
- Grammar
- Vocabulary
- Fluency
- Other

18.

Why do you choose to assess the learners on this?

19.

If you do not assess learners on their use of English, please explain why:

20.

Have you used one (or more) of these exercises in your tests this year? If yes, please tick the box:

- gap text
- put words in the right order
- circle correct answer
- reading & understanding
- creative writing
- argumentative/descriptive essays
- vocabulary
- listening
- presentations
- summarizing
- other ...

21.

Please list the exercises you use most and describe them in more detail, including your reasons for using these specific exercises:*

The following questions are about the admission procedure/test.

22.

All learners in bilingual education at the HWC went through the same admission procedure. Are you aware of the results they have obtained?*

- Yes, for all students
- Yes, for most students
- Yes, for some students
- No

23.

If yes, do you use this knowledge during your lessons? (i.e. challenge or help pupils based on their results) How?

If no, would you like to know the results and/or see the test? Please explain why:*

24.

May there be any questions, can we contact you? If yes, please fill out your email address below: Also, feel free to add any comments in the box below.

Thank you!

APPENDIX C – ANSWERS TO QUESTIONNAIRE

Respondent 1: history

Respondent 2: biology

Respondent 3: history

Respondent 4: geography

6. Have you been teaching according to CLIL philosophy? If so, since when?

Res1: Since the beginning although I have to remind myself to integrate CLIL in my lessons

Res2: Since start bilingual stream

Res3: Yes, even since my educational master at the University of Utrecht. I also followed an in-depth course on CLIL after I graduated.

Res4: Since the start of the bilingual stream

7. What is your definition of CLIL and how do you apply CLIL in your classroom?

Res1: The integration of the English lessons in my history lessons. I try to use CLIL exercises to do so. I'm afraid I don't always have time to do this but just reading together, making word webs (placemats), doing exercises (odd one out, watch English videos, read articles) are things we do regularly

Res2: e.g. Project endangered species not only in biology class, but also in English class. Both teachers present at presentations. Mark for content biology, Mark for English (grammar, spelling etc.)

Res3: Content and Language Integrated Learning? Yes, I practice CLIL through many different learning activities. These different activities are aimed at creating different forms of language output. Many activities can be found in literature or online, although I occasionally try to devise my own CLIL-assignments

Res4: Yes.

8. + 9. How do you experience CLIL in terms of English acquisition, in general (i.e. at other schools, previous jobs)?

Res1: Neutral, why: I don't really understand the question. I can only answer this for my students and for my lessons.

Res2: Excellent, why: skills students need to develop taught in both classes is an advantage

Res3: Good, why: CLIL is a mandatory and certainly necessary instrument for any TTO-teacher (many CLIL activities are also perfectly applicable to Dutch-spoken lessons). As a subject teacher however, the primary goal of my lessons is to convey historical knowledge and skills. Language acquisition is an important, but secondary (or auxiliary) goal within my lessons.

Res4: Good.

10. + 11. What is your opinion on CLIL?

Res1: Good, why: I think CLIL helps the students acquire English and they have a lot of knowledge about the English lessons, even after a couple of months

Res2: Excellent, why: Students experience a subject teacher is a language teacher As well

Res3: Good, why: as stated above, the main focus of my lessons is History. Naturally, a proper understanding of the language is a prerequisite for learning and I try to include at least 1 'CLIL-ish' activity in each lesson (80 minutes). However, my priority is teaching my own subject (which can be done through CLIL, next to other strategies).

Res4: Good.

12. + 13. Which English skills are most important during your lessons?

Res1: reading 4, writing 3, listening 5, speaking 5 - Why: I do not correct the grammar/spelling of the students. But I do focus on encouraging speaking and listening

Res2: reading 3, writing 3, listening 3, speaking 3 - Why: I think students need all the skills. Related to learning strategies

Res3: reading 4, writing 2, listening 5, speaking 4 - Why: Reading and listening are the (most) elementary skills for TTO students, especially in the first form.

Res4: reading 5, writing 3, listening 5, speaking 4.

14. + 15. Do you mark language mistakes in learners' work?

Res1: No, I teach history and not English. If I teach in Dutch I don't mark mistakes either. I do underline mistakes but I'm not distracting points

Res2: Yes, I think it is important

Res3: Sometimes, It depends on the classes (forms) and assignments

Res4: No.

16. - 19. Are learners assessed on their English in your tests?

Res1: No, I teach history and not English. If I teach in Dutch I don't mark mistakes either. I do underline mistakes but I'm not distracting points

Res2: No, I do not have the time for it. Common problems or mistakes I discuss in class.

Res3: Sometimes, on Reading comprehension, Listening comprehension, Speaking, Writing, Grammar, Vocabulary. Why?: because it's necessary for their language skills to develop, this in turn enables the students to have a better understanding of my subject.

Res4: No.

20. Have you used one (or more) of these exercises in your tests this year? If yes, please tick the box:

Gap text	(R1, R3, R4)
Put words in the right order	(R1, R2)
Circle correct answer	(R2, R3, R4)
Reading & understanding	(R1, R2, R3, R4)
Creative writing	(-)

Foretelling or Fortuitous?

Argumentative/descriptive essays	(R1)
Vocabulary	(R2, R3, R4)
Listening	(-)
Presentations	(R1, R2, R3, R4)
Summarizing	(R1)

APPENDIX D – DATA

Fig. 1.0 Correlation

This Figure shows us a scatterplot created from the admission test outcomes and the Geography grades from Term 1, to test the reliability of the Standard Correlation test. There are no (extreme) outliers, which means that the data for this test is practicable.

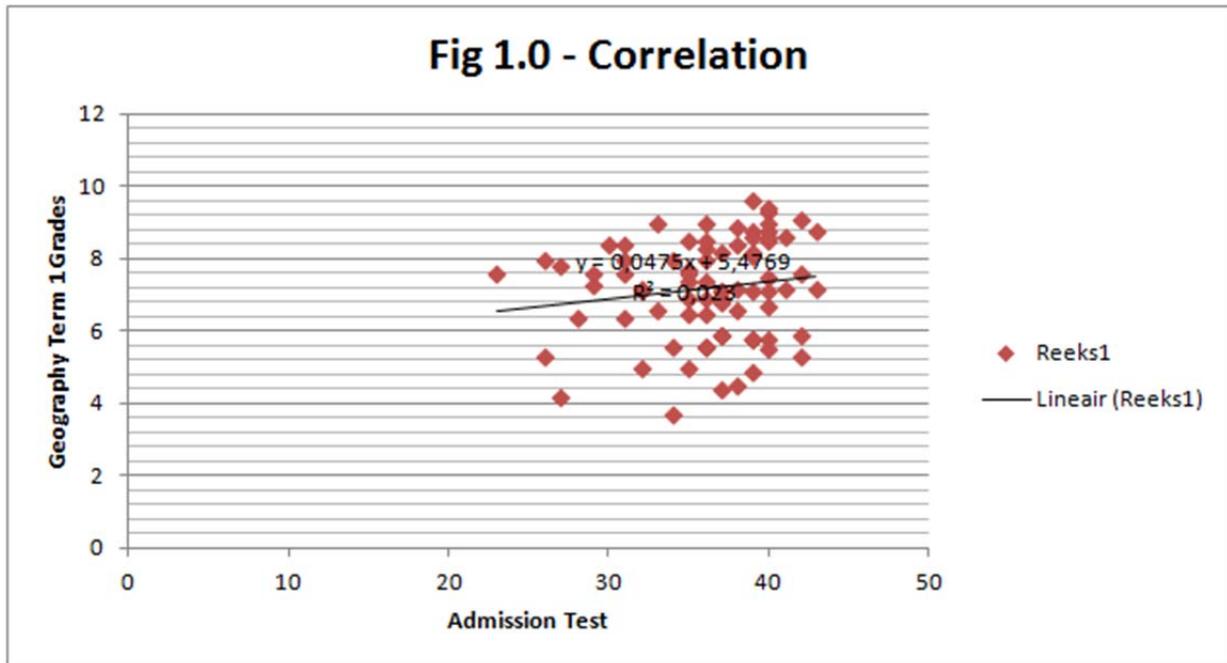


Fig. 1.0. $y=0.0475x + 5.4769$. $R^2 = 0.023$.

Table 1: Independent T-Test

This table shows us *Levene's Test of Equality of Variances* and *Sig. (2-tailed)*, which allows us to read the significance of the admission test outcomes compared to the term grades for each group.

R1 stands for Term 1, R2 for Term 2, R3 for Term 3 and 'hi' stands for History, 'ge' for Geography and 'bg' for Biology. So, 'hiR1' means History, Term 1 grades.

Independent Samples Test						
		Levene's Test for Equality of Variances		t	df	Sig. (2-tailed)
		F	Sig.			
hiR1(1)	Equal variances assumed	1,311	,270	-1,644	15	,121
	Equal variances not assumed			-1,709	11,828	,114
hiR2(1)	Equal variances assumed	,677	,423	-.964	15	,350
	Equal variances not assumed			-.991	13,556	,339
hiR3(1)	Equal variances assumed	,001	,972	-1,495	15	,156
	Equal variances not assumed			-1,513	14,915	,151
geR1(1)	Equal variances assumed	,013	,912	-.757	15	,461
	Equal variances not assumed			-.758	14,831	,461
geR2(1)	Equal variances assumed	1,165	,297	-.909	15	,378
	Equal variances not assumed			-.926	14,592	,370
geR3(1)	Equal variances assumed	,761	,397	-1,157	15	,265
	Equal variances not assumed			-1,180	14,480	,257
bgR1(1)	Equal variances assumed	,668	,426	-.665	15	,516
	Equal variances not assumed			-.658	13,861	,521
bgR2(1)	Equal variances assumed	,679	,423	-.926	15	,369
	Equal variances not assumed			-.936	14,968	,364
bgR3(1)	Equal variances assumed	,540	,474	-1,109	15	,285
	Equal variances not assumed			-1,124	14,847	,279

Table 1. The Significance of the Independent Samples T-Test; Sig. (2-tailed).

Fig 2.0 Regression

To make sure the results were valid, we looked at the linear relationship between the two variables, checking for significant outliers to prevent a negative effect on the regression equation, and the independence of observations (Laerd Statistics 2013). An example of this can be seen in *Fig 2.0*. This shows us a scatter plot created from the admission test outcomes and the History grades from Term 1, to test the reliability of the regression analysis.

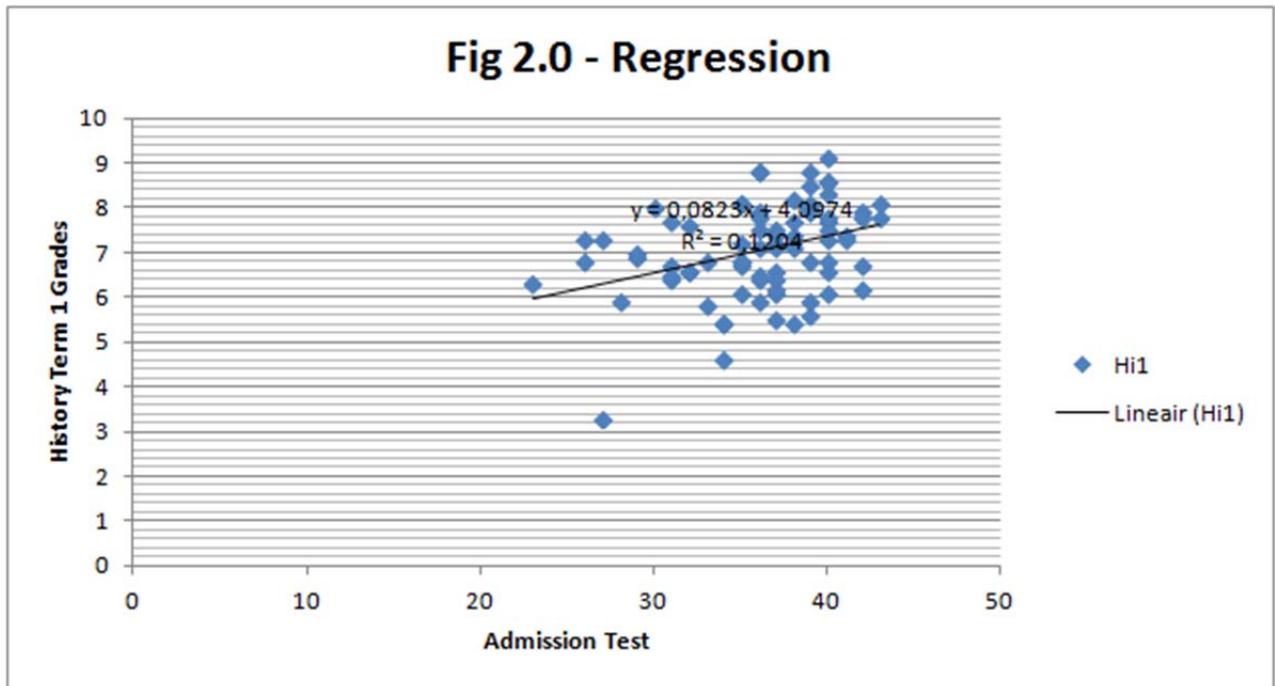


Fig. 2.0. $y=0.0823x + 4.0974$. $R^2 = 0.1204$.

Fig 3.0. Model Summary, ANOVA

The Model Summary table, shows the R value that represents a simple correlation, and the R² value, which “indicates how much of the dependent variable (in this case CLIL subject term grade) can be explained [or predicted] by the independent variable (in this case admission test outcomes)” (Laerd Statistics 2013).

The ANOVA table indicates whether the regression model predicts the outcome variable significantly well. This can be seen in the regression row and Sig. column. If the number is less than 0.05, it means that the applied model can “statistically significantly predict the outcome variable” (Laerd Statistics 2013).

Intake stands for admission test outcomes, ‘hiR1’ stands for History Term 1.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	intake, egR1 (1) ^b	.	Enter

a. Dependent Variable: hiR1(1)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,498 ^a	,248	,228	,9104

a. Predictors: (Constant), intake, egR1(1)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20,802	2	10,401	12,548	,000 ^b
	Residual	62,993	76	,829		
	Total	83,794	78			

a. Dependent Variable: hiR1(1)

b. Predictors: (Constant), intake, egR1(1)

Fig. 3.0.. Model Summary & ANOVA.

Fig. 4.0. Coefficients Table

The Coefficients Table “gives us the information we need to predict [a CLIL subject term grade] from [the admission test]” (Laerd Statistics 2013). If the number in the Sig. column and the admission test (“intake”) row is $p < 0.05$, we can say whether the admission test contributes significantly to the CLIL subject term grade.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,315	,884		3,748	,000
	egR1(1)	,369	,103	,427	3,595	,001
	intake	,027	,028	,113	,953	,344

a. Dependent Variable: hiR1 (1)

Fig. 4.0.. Coefficients Table