

The CEFR as an Effective Tool for Evaluation Used by Secondary School Pupils

By Marvin Willems

marvin.willems@student.ru.nl

BA Thesis English Language and Culture

Supervisor: Dr. C.M. de Vries

15 June 2016

ENGELSE TAAL EN CULTUUR

Teacher who will receive this document: **Dr. C.M. De Vries**

Title of document: **Vries\_Willems\_BAthesis\_June**

Name of course: **BA Thesis**

Date of submission: **15 June 2016**

The work submitted here is the sole responsibility of the undersigned, who has neither committed plagiarism nor colluded in its production.

Signed

Name of student: **Marvin Willems**

Student number: **4141601**

### **Abstract**

The goal of this research was to establish whether pupils in secondary schools are able to use the Common European Framework for Reference (CEFR) effectively as a tool for evaluation. Which means that they can use the CEFR to assign the right level of proficiency to, in this case, a text. This thesis looked at secondary school pupils in year one, three and five of their pre-university education in the Netherlands. Their use of the CEFR as a tool for evaluation has been explored through an experiment in which the pupils were asked to assign a CEFR level to two texts in the categories vocabulary, punctuation, and coherence. Results showed that many pupils were not able to assign the right level, but that their scores were increasing as the pupils grew older. Another observation was that vocabulary was the most difficult category to evaluate and that standard deviations become smaller as the pupils get older, meaning a smaller spread of answers across all answer options. In other words, the answers seem to centre more on one point as the pupils get older.

Keywords: CEFR, descriptors, education, English, evaluation, levels

**Contents**

Abstract .....	2
1. Introduction .....	4
2. Background .....	6
2.1 What is the Common European Framework for Reference? .....	6
2.2 Different uses of the CEFR.....	7
2.3 Opinions regarding the CEFR .....	8
2.4 Basis for this thesis .....	11
3. Methodology .....	12
3.1 The participants .....	12
3.2 The task.....	13
3.3 The setting .....	14
3.4 Analysis .....	14
3.5 Predicted results.....	15
4.1 The overall levels in text A and B.....	16
4.2 The descriptors of texts A and B per year.....	17
4.3 Normality per category and year .....	18
5. Discussion .....	20
5.1 The overall levels of text A and B .....	20
5.2 The difficulty of the descriptors .....	21
5.3 Normality per category and year .....	22
6. Conclusion.....	24
6.1 Flaws and future research .....	26
7. References .....	28
8. Appendices .....	30
A - Experiment text A .....	30
B - Experiment text B .....	31
C - Complete experiment.....	32

## 1. Introduction

The Common European Framework for Reference (CEFR) is a framework that is used in a lot of countries in Europe and is designed to give insights in the different levels that a language learner can attain. These levels range from A1 to C2 and contain many descriptors in several different categories such as vocabulary range, grammar, coherence, and punctuation. The CEFR is used in the Netherlands as a guideline in language curricula design and in the design of national exams. The CEFR is also often used by teachers who use it as a guideline to evaluate their pupils.

In my experience as a secondary school teacher I found that the CEFR can be a useful tool for pupils to provide insight in their current level. Thinking back to my years as a pupil I remember that I did not quite understand that the CEFR could be useful for me as, nor did I understand that the CEFR had such an impact on my evaluation and the final exams in my secondary school. Now I see that it is in fact a useful tool, but I feel that it needs to be improved upon in order to be more transparent towards pupils so that they can use the CEFR effectively. The CEFR should be more transparent towards pupils, especially because they are often evaluated according to the framework and with the CEFR being more transparent, pupils get a better insight in their current level and will gain a better understanding of what they need to do in order to improve themselves.

The goal of this thesis is to ascertain whether or not pupils can use the CEFR effectively as a tool for evaluation. This will be explored through an experiment in which pupils are asked to assign a certain CEFR level to a text. The categories they must evaluate are vocabulary, punctuation, and coherence. The focus of this research is on the pupils in secondary school. More specifically, pupils currently in year one, year three and year five of pre-university education.

This thesis is structured as follows. Chapter two will provide relevant background information as well as a theoretical framework which leads to the research question or main hypothesis of this thesis. The methodology used in this thesis and the design of the experiment will be discussed in chapter three ending with the sub-hypotheses which will help to formulate an answer to the main hypothesis. The main hypothesis is whether pupils in secondary school can use the CEFR effectively as a tool for evaluation and to answer this main hypothesis, sub-hypotheses have been constructed. The first set of sub-hypotheses are looking whether pupils are able to assign the correct overall level for two texts and the second set of sub-hypotheses will seek to find an answer to which category (vocabulary, punctuation, or coherence) is most difficult to evaluate. Chapter four will give an overview of the data and

the analyses that are used and chapter five will discuss those results. Chapter six will provide a conclusion to this thesis.

## **2. Background**

This chapter will provide relevant background information about the CEFR, its uses, and any issues around the CEFR. These topics encompass the literary theory that will be the groundwork on which this thesis was built. The end of this chapter will zoom in on the particular issue that is at the foundation of this thesis.

### **2.1 What is the Common European Framework for Reference?**

Second language learning can be a difficult process and with the globalisation at hand second languages have become a vital part of our society. In the Netherlands, English is often taught in the final years of primary school and is continued at secondary school where even more languages can be studied such as French, Spanish, and German. The perspective on the learning of second languages has changed significantly over the years. In the first half of the 20<sup>th</sup> century, foreign languages were primarily taught through grammar translation. The idea then was that being able to read and write a foreign language as well as their native language is sufficient to communicate with others (Larsen-Freeman & Freeman, 2008). Which means that there was hardly any attention being paid to speaking and listening skills. Nowadays, the perspective has changed; it is now more important to focus on the learners' ability to use a language effectively in various circumstances. To use a language effectively it is important that the pupils do not only learn how to use grammar correctly, but also how they need to write, read, listen, and speak in this foreign language. This new focus on language learning led to a need for proficiency scales that could be used in language learning contexts (Larsen-Freeman & Freeman, 2008).

This change in focus is one of the main reasons that led to the establishment of the Common European Framework for Reference for languages (CEFR). The CEFR is a practical tool and helps teachers and learners to value even the smallest achievements in second language learning (North, 2014). It is devised as a framework and meant only for reference and is therefore intentionally not language specific. This means that the CEFR should be used as a guideline rather than a strict set of rules to which every European country must adhere to. The CEFR supports comparative mapping of competences in several languages in order to find a set of standards. The CEFR is a framework developed for learning, teaching and assessing. The concept of the CEFR rests on two very closely related pillars; quantity and quality. The descriptions are on a wide range of subjects such as grammar and vocabulary control, and indicate how well the learner performs in particular areas (Hulstijn, 2007). It is a framework that comprises a descriptive scheme for analysing what is needed in language learning and use, as well as a definition of communicative proficiency at six levels arranged

in three categories; A1 and A2 (basic user), B1 and B2 (independent user), and C1 and C2 (proficient user) (Little, 2007). In other words, it does not only say what one needs to do to acquire a certain level, it also gives an indication on how well the learner is expected to use language at any given level. The descriptive scheme can be considered to be, in Mislevy's terms, a learner model, "... a simplified description of selected aspects of the infinite varieties of skills and knowledge that characterise real students" (Mislevy in North, 2007b, p.23). Krumm describes the CEFR as being extremely useful and influential, and it will most likely continue to have this role in language learning contexts (2007). As Alderson stated "Nobody engaged in language education in Europe can ignore the CEFR. The CEFR itself has contributed to more transparent curriculum designs and examinations in several European nations" (Alderson, 2007, p.660).

CEFR is often used in educational contexts, and language learners, language teachers, and employers find the framework very helpful in setting the curricular system in their own countries and regions, or when hiring employees. "The CEFR is strictly neutral as regards teaching or testing methodologies. It is intended to be 'comprehensive' in that it is possible to situate any style of teaching within the conceptual framework provided" (North, 2007b, p.27). The framework also helps people to be more specific about what students can and cannot do in the language they are acquiring (Hulstijn, 2007). The CEFR was initiated as a response to the changing views on language learning, but it is said that the CEFR will cause language education to change. Some say that the CEFR has not changed a thing in curriculum planning (Little, 2007), but others argue the opposite and say that every change in second language education happens at a very slow pace (Trim in North 2007b).

The descriptors in the CEFR are written from the learner's point of view. This allows the learner to compare what he can do with what he wants to be able to do and thus encourages the learner to improve their skills (Larsen-Freeman & Freeman, 2008). This would lead to better results, because language learners know more specifically what their level is and what they can achieve. Having a specific goal is a way that can motivate language learner to continue studying a language.

## **2.2 Different uses of the CEFR**

The CEFR can be used in any language learning environment and has a lot of influence in Europe with regards to curricula design, learner and test assessment, and mapping language learning (Krumm, 2007). The CEFR is used as a guideline for the level of exams in several countries such as the Netherlands. It is not obligatory to use it, but the CITO (a Dutch

national exam committee) established a specialised task force that focuses on how the exams line up with the CEFR scales (Moonen et al., 2010). The CEFR was intended as a tool that would enable language learners to say where they were at a certain point in time, not a description that tells language learners where they ought to be (North, 2007b). It is also used as a guideline for language learning curricula by many schools in Europe (Moonen et al., 2010). This does raise a few questions, because the authors of the CEFR were not very explicit about the implications for classroom teaching, nor was the CEFR inspired on solid second language acquisition research. In other words, the descriptors tell us what language learners can do at a certain level, but it does not describe what the learners need to know in order to carry out several language tasks (Westhoff, 2007). The CEFR can be a useful tool in classroom settings, but teachers have to teach their pupils how to use it effectively, which might not be the case now.

### **2.3 Opinions regarding the CEFR**

The CEFR is very useful tool and many people are very enthusiastic about the framework. There are also people who think that the CEFR can be improved so that it fulfils the needs of more language learners around the world. Previous research has shown that the CEFR is a good framework for language teachers and learners, but it can be improved in order to create an even better framework. Some say it is far from perfect and needs more improvement whilst others say, including CEFR itself, that this is a good framework in itself and should be used as a guideline rather than as strict scales. Hulstijn claims that “[the CEFR] is, of course, not perfect, but it is good enough to be improved upon and developed further” (Hulstijn, 2007, p.666).

Larsen-Freeman and Freeman do have some issues regarding the CEFR, but they also argue that there are a lot of things that the CEFR has, which makes the CEFR as useful tool for reference. The CEFR being a descriptive framework that is not language specific is a huge step forward, because it seeks to distinguish from a set of standards, which was the case in the United States (2008). Another aspect that is highlighted in this article is that the CEFR descriptors are largely presented as can-do statements and are presented from the learner’s point of view. These statements are often written in a positive way, saying what a language learner can do, rather than saying something what the language learner cannot do yet. This causes the language learner to compare that what he is able to do with what he wants to be able to do (Larsen-Freeman & Freeman 2008). This may encourage a language learner to go more far in studying a language until it meets the needs for that specific language learner. A

language learner who is learning a new language to communicate with friends might feel that his needs are sooner fulfilled than a language learner who needs to maintain business contacts abroad.

There are many arguments to be found against the CEFR. One of which is that the levels are linear (Krumm, 2007). He goes on to say that pupils have to show the same skills or are expected to show the same skills in reading, listening, writing, and speaking in order to be evaluated on a B2 level for example. This implies that every pupil follows the same route when it comes to language learning. The CEFR can be understood as a framework that expects language learners to move through level A1, A2 and B1 before they eventually arrive at level B2. Although Krumm is right to a certain extent, it is not the CEFR who is to blame for this, because the exams and methodologies that are made -some of which claim to be based on the CEFR- are often made linear in order to work with an entire class in the same pace without much differences amongst the pupils. Exam institutions create their exams also in a linear fashion and those are not set up by the CEFR itself, but by people who used the CEFR merely as a guideline. Krumm also states that one has to show A2 proficiency in all the different areas -speaking, listening, writing, and reading- in order to be scaled as A2(2007). Whilst this may seem true, this is not the case for every test. There are also tests in which it is possible for a learner to exceed in one or two of these areas, but one could say that as a more general rule the level that is apparent in most areas is the overall level. The CEFR can thus be used as a guide to help develop tests and its descriptors will help the teacher being more objective (North, 2014).

Another aspect which is frequently commented upon are the differences between learners in motivational, attitudinal, and aptitude factors (Larsen-Freeman & Freeman 2008). Language learners with a high level of motivation are often more prepared to go the extra mile when things tend to go wrong or when they have a specific goal or reason for studying a language. Whereas language learners with a low level of motivation are often not willing to put the extra time and effort in learning a language. The same can be said for attitudinal and aptitude factors which can influence an individual significantly. Larsen-Freeman and Freeman's comment would then mean that a framework is not the right tool at all, because it does not take individual factors into account. Although it is true that every learner has his own path in acquiring a new language, the framework is designed not to be specific for a reason. The CEFR is applicable to many languages and for many language learners, because of its lack of specificity. This is what makes the CEFR able to be used in many different nations with many different language learners, tests, and curricula.

Apart from being not specific enough, it is often said that another main issue is that the CEFR is not validated by empirical research. The descriptors for written production were mainly developed from those for spoken production (Little, 2007). To take it even further, the CEFR is also not based on second language acquisition (SLA) research. The CEFR is designed by teachers' perceptions of language proficiency (North, 2007a). This criterion is also shared by North who says that "at the time that CEFR was developed, SLA was not in such a position to provide these descriptions" (2007a). Hulstijn agrees with both Little and North on the lack of empirical evidence and argues that it is high time for the CEFR to re-evaluate its framework and engage in collaborative research with SLA researchers (Hulstijn, 2007). Re-designing the CEFR with the help of second language acquisition research might improve the quality of the CEFR and would contribute to the CEFR's validity.

Charles Alderson (2007) raises some more issues regarding the CEFR. Firstly, he finds that the CEFR is limited because it is not language-specific. Secondly, he states that many translations of the original CEFR are much clearer than the original and that the CEFR is vaguely defined and therefore not reader-friendly. The Dutch CEFR Construct Project investigated the usefulness of the CEFR for testing and found that "many terms lacked definitions, there were overlaps, ambiguities, and inconsistency in the use of terminology, as well as important gaps in the CEFR scales" (Alderson, 2007, p.661). Those two problems combined make that the CEFR in its present form is highly unsuitable for young language learners (Alderson, 2007). In response to this Figueras says that the CEFR is never meant to be used by young language learners, but by adult foreign language learners (Figueras, 2007). However, a project has been started in the area of Languages of Education in order to develop CEFR descriptors that are suitable for young learners in various disciplines and genres (North, 2014).

Another issue which remains unresolved is how the quality of the tests is controlled. Who knows for certain that a test of Greek calibrated at level B1 in Finland is equivalent to a test of Polish considered to be at level B1 in Portugal? Bonnet continues to argue that the CEFR should not only be improved with regard to the scales and descriptors, but the teachers and language coaches should also be taught how to use the CEFR effectively (Bonnet 2007). This is an interesting point and can only be answered with an extensive research concerning this issue. It might very well be that different tests which should be at the same level are not.

## **2.4 Basis for this thesis**

Many of these researchers found that the CEFR descriptors are vague and that the CEFR should not be used by young language learners. However, the CEFR is frequently used in secondary schools in the Netherlands and because of that it is important to see whether pupils know how to evaluate themselves based on the CEFR scales, and, more importantly, if they fail to do so, what it is exactly that they find difficult. This paper expects to come up with a diagnosis on the question what the effect is of the use of the CEFR as a tool for evaluation by pupils within classroom environments in secondary schools. This could then be used as a starting point for further research on how to make the CEFR clearer for pupils or, if the results are all very positive, it will bring validity to the use of the CEFR in classroom environments in secondary schools.

### **3. Methodology**

This chapter will describe the design of the experiment that is used to answer the research question and the hypotheses mentioned in section 3.5. The participants and the task will be described, as well as the manner in which the data were collected and analysed with the help of the computer programme *IBM SPSS*.

This experiment will aim to find out whether pupils are able to evaluate pieces of writing based on the descriptors on writing as drawn up by the CEFR. These descriptors were the main source for creating the descriptors in Dutch that will be used to evaluate the pieces of writing. These descriptors are translated to Dutch by the SLO, a Dutch national institute for curriculum development in the Netherlands. In short, the pupils will be asked to evaluate two pieces of writing based on the CEFR and to validate their evaluation by giving arguments and examples from the text. Every pupil will get the same two pieces of writing which need to be evaluated. The data will show whether pupils are able to evaluate a piece of writing correctly according to the CEFR scales or not. The data will also show what pupils find difficult in aligning work with the CEFR descriptors, because the pupils need to give arguments supporting their claim. The pupils need to evaluate a piece of writing and have to take a critical look in order to find the correct level. Due to the scope of this thesis it is assumed that if the pupils are able to evaluate work which is written by others, they are also able to evaluate their own work by looking at it critically. This critical look is the base for the assumption that pupils can look critically at their own work too, but this assumption, will need to be researched in the future to see whether this is true or not,

#### **3.1 The participants**

The participants in this experiment are three groups of pupils from a secondary school. Each group consisted of around twenty-five pupils. The reason for choosing pupils in secondary schools is they are often evaluated by using the CEFR and many (nation-wide) test results in the Netherlands are aligned with the CEFR as well. The participants in this study are all attending pre-university education in secondary school. Pre-university education is called VWO in the Netherlands and pre-university education will thus now be referred to as VWO in the rest of this paper. The participants were all pupils from the Almende College location Isala in Silvolde. The first group consists of first-year pupils from ages twelve to thirteen, the second group consists of third-year pupils with ages ranging from fourteen to fifteen, and the third group consists of fifth-year pupils with ages ranging from sixteen to eighteen. The pupils in this secondary school are familiar with the CEFR. First year pupils know that some writing assignments are evaluated by the CEFR and third and fifth year pupils know that their writing

assignments and speaking tests are evaluated by the CEFR. The third and fifth year pupils also used a simplified before which was drawn up by their teachers and which is inspired on the CEFR. These three different groups are chosen because they may show different results in their capability of evaluating according the CEFR descriptors, because fifth-year pupils are more likely to have more experience with using the CEFR as well as having more experience in reading and writing English than first-year pupils.

### **3.2 The task**

The pupils are asked to evaluate two pieces of writing. The piece “How Many Ways Do I Love Thee” is collected from the website of the CEFR (Europees Referentiekader Talen, n.d.) and is assigned to a CEFR level by the CEFR itself. The CEFR evaluated some texts and put them online in order to give an example of how work can be evaluated according to the CEFR descriptors. The second piece of writing is collected from the personal database of Dr. De Vries and is evaluated by students and teachers according to the same descriptors as the first text. The first text “How Many Ways Do I Love Thee?” (ERK n.d.) (Appendix A), is rated as an A2 level by the CEFR and the second text “A Teacher’s Profession Is an Ideal Profession. Yet All Teachers Are Not Ideal.” (Stevens, 2015) (Appendix B), is rated as a B1 level by university students and secondary school teachers. Pupils are given the pieces of work that they need to evaluate based on the CEFR. The pupils are also given a scheme with three categories, vocabulary range, spelling, punctuation, lay-out, and coherence. Pupils are asked to assign a CEFR level to each of the three categories and are asked to provide arguments or examples as well. The descriptors provided by the SLO are in Dutch in order to prevent any misconceptions regarding the descriptors, especially for the younger pupils who are less fluent in English. Some of the descriptors contained some difficult words for the first year pupils, so these words are given a synonym to clarify the descriptor, but the descriptors are not more simplified in order to maintain the original CEFR structure. The pupils fill out this scheme according to their individual beliefs and are also asked to give arguments or examples wherever possible.

The task will start with the text that was submitted and rated by the CEFR. After each text is a scheme with the different categories and the CEFR levels regarding writing. Every category has a descriptor for each level (A1-C2). Pupils are supposed to choose one level per category. This is repeated for the text “A Teacher’s Profession Is an Ideal Profession. Yet All Teachers Are Not Ideal.” (Stevens, 2015). During the task the pupils are not allowed to talk to

each other in order to prevent them from getting influenced by others and to limit the variables as much as possible. An example of the experiment can be found in appendix C.

### **3.3 The setting**

The pupils received the task in a classroom setting and only needed to use a pen or pencil. The task itself was printed and was to be handed out in the classroom. Each pupil received one copy of the task on which they could fill out the scheme and answer the questions. The task was anonymous and pupils were only asked to give their age, gender, and in which class they were. There was no need for any other facilities such as laptops, iPads, or dictionaries. The instructor briefly explained how this task works in Dutch to prevent any confusion. After the instruction the pupils could commence with their tasks. The instructions were also given in steps on the task itself so that the pupils could always find out what they needed to do next. During this experiment the pupils were not allowed to confer with one another so as to prevent them from being influenced by each other. The task took around twenty minutes at most and the pupils were asked to read two different texts and evaluate both texts according to the descriptors. The texts were written in English, but the descriptors were translated into Dutch. When the pupils had finished their tasks, the instructor collected the papers.

### **3.4 Analysis**

The dependent variables in this experiment are the columns gender, vocabulary, punctuation, coherence. The independent variables in this experiment are the classes and the texts provided. *IBM SPSS* is a software package used for statistical analysis (Field, 2013). Statistical analyses are used to find significant differences between groups. If something is significant that means that it is statistically likely that a difference is rightly observed rather than being incidental. It is possible that the results are not representative of the wider population, in this case of the Netherlands. A cut off point is used to set a boundary. This cut off point is usually 5%. This means that a one in twenty chance is taken that a result is representative of the population. The statistical analyses which will be used are Anovas and paired sample t-tests. An Anova looks at the averages of groups and shows whether there is a significant difference. Apart from using quantitative methods, qualitative methods will be used as well in order to gain a better insight in the results. The different categories vocabulary, punctuation, and coherence for text A are referred to as follows A\_Vocab, A\_Int, and A\_Coh. The same notation is used for the categories in text B.

### 3.5 Predicted results

Once all the results are in, they will be analysed for different means. The results will be analysed in such a way that the overall levels that are assigned to the texts are compared between years. The expectation is that the pupils in their fifth year are better at assigning the right level than the first and the third year pupils. In this context, better means that pupils connect the right levels with the right texts. This would imply that more experienced learners of English are more capable of using the CEFR successfully. This is expected for both texts and may have to do with the fact that they have more experience with the CEFR and that they are more experienced with reading and writing English. It is also expected that third years are also better at assigning the correct levels to the texts for the same reasons as mentioned earlier.

Another expected result has to do with the CEFR descriptors. It may be possible that it is easier to assign a certain descriptor more successfully than another. The descriptors here are vocabulary, punctuation, and coherence, and from personal experiences I expect that vocabulary is easier to assign than the other two, because vocabulary is a topic which is paid a lot of attention to from the first year. This expected result assumes that the difficulty of the different CEFR descriptors does not depend on the difficulty of the texts.

All of these analyses will help in answering the main hypothesis on whether the CEFR can be used effectively by pupils in secondary schools in the Netherlands. The expected results are more clearly defined in the following hypotheses.

H<sub>1a</sub>: Year 5 is better at connecting the right overall level to a text than year 1. (text A)

H<sub>1b</sub>: Year 5 is better at connecting the right overall level to a text than year 3. (text A)

H<sub>1c</sub>: Year 3 is better at connecting the right overall level to a text than year 1 (text A)

H<sub>2a</sub>: Year 5 is better at connecting the right overall levels to a text than year 1 (text B)

H<sub>2b</sub>: Year 5 is better at connecting the right overall level to a text than year 3. (text B)

H<sub>2c</sub>: Year 3 is better at connecting the right overall level to a text than year 1 (text B)

H<sub>3a</sub>: X\_Vocab is easier to judge than X\_Coh (texts A+B)

H<sub>3b</sub>: X\_Vocab is easier to judge than X\_Int (texts A+B)

H<sub>3c</sub>: X\_Int is easier to judge than X\_Coh (texts A+B)

## 4. Results

This chapter will present all the results that were collected. The data was analysed with the help of *IBM SPSS* version 19. The data was analysed with aid of quantitative methods as well as with the aid of qualitative methods. The quantitative methods used to analyse these data were Levene's tests, Shapiro-Wilk tests, Anova tests and paired sample t-tests.

After conducting the experiment, the pupils were asked whether they thought this experiment was difficult or easy. Most of them replied that it was quite difficult to find the right descriptors and the examples from the text to support it. This chapter will merely present the results that this experiment yielded. A discussion of these results is to be found in the next chapter. The results of the overall levels assigned to text A and text B will be presented first. Second, the results concerning the descriptor categories are shown, and finally the results concerning normality are presented with the aid of two tables showing an overview of the standard deviations per category and year.

The data that was collected were put into *IBM SPSS*, and these will be compared using Anova tests and paired sample t-tests. The Anova tests were performed with Year as a fixed factor and  $\alpha$  was set on 0.1 every time. This showed whether the right levels were assigned or not, and whether the data are significant. The paired sample t-tests were performed between the descriptors (vocabulary, punctuation, and coherence) that are used for the A and B texts. This will show whether the pupils from different years were better at assigning the right level in a certain category than in another, i.e. whether first years had a better notion of how to evaluate vocabulary than third year pupils, for example. Levene's tests of equality of variances need to be performed before an Anova or paired sample t-test can be executed as these tests only work under the assumption that the variability in the two groups or conditions is not significantly different. The data also have to pass the assumption of normality. It is assumed that data are normally distributed when this assumption is passed. This was tested using the Shapiro-Wilk test.

### 4.1 The overall levels in text A and B

An Anova test was performed with Year as a fixed factor and A\_Vocab, A\_Int, A\_Coh, and A\_Overall as dependent factors. Shapiro-Wilk tests are significant for all the dependent variables. Levene's test was significant for A\_Overall ( $p < 0,05$ ). Levene's test was not significant for A\_Vocab, A\_Int, and A\_Coh. The target M for text A is 2 in the overall level as well as the separate categories. A significant result was found for A\_Vocab ( $F = 2.329$ ,  $p = 0.052$ ). A contrast test for A\_Vocab showed that fifth year pupils ( $M = 3.04$ ) performed

significantly better ( $p = 0.046$ ) than first year pupils ( $M = 3.54$ ). An almost significant result was found for A\_Coh ( $F = 1.551, p = 0.1085$ ). A contrast test for A\_Coh showed that fifth year pupils ( $M = 2.43$ ) performed better ( $p = 0.087$ ) than first year pupils ( $M = 2.93$ ) and although this result is not significant, it shows a contrast between the pupils.

An Anova test was performed with Year as a fixed factor and B\_Vocab, B\_Int, B\_Coh, and B\_Overall as dependent factors. Shapiro-Wilk tests are significant for all the dependent variables. Levene's test was significant for B\_Coh ( $p < 0,05$ ). Levene's test was not significant for B\_Vocab, B\_Int, and B\_Overall. No significant results were found in the Anova test for B\_Vocab ( $F = 1.961, p = 0.148$ ), B\_Int ( $F = 1.548, p = 0.637$ ), and B\_Overall ( $F = 0.029, p = 0.971$ ). In conclusion, text B did not yield any significant results.

This quantitative analysis seems insufficient, since language is dealt with. Therefore, a qualitative analysis seems more appropriate in this case. The overall levels which were correct are analysed per year and put in a table. The columns 'Text A' and 'Text B' show only the absolute number of the correct levels for texts A and B per year, while 'Text A %' and 'Text B %' show how many percent per year had the correct overall level.

**Table 1. The number of pupils who assigned the right level and the percentages per year.**

<b>Year</b>	<b><i>n</i></b>	<b>Text A</b>	<b>Text B</b>	<b>Text A %</b>	<b>Text B %</b>
<b>1</b>	28	5	5	17.9%	17.9%
<b>3</b>	25	6	2	24%	8%
<b>5</b>	28	12	1	42.9%	3.8%

#### **4.2 The descriptors of texts A and B per year**

The first years' results showed some significant results in the descriptors of text A. A\_Vocab paired with A\_Int was almost significant ( $t(27) = 1.544, p = 0.67$  (1-tailed)). A\_Vocab paired with A\_Coh showed a significant result ( $t(27) = 2.684, p = 0.006$  (1-tailed)). A\_Int paired with A\_Coh showed no significant result ( $t(27) = 1.070, p = 0.147$  (1-tailed)).

The first years' results for text B showed some significance regarding the descriptors. B\_Vocab paired with B\_Int showed a significant result ( $t(27) = 2.274, p = 0.0155$  (1-tailed)). B\_Vocab paired with B\_Coh showed a significant result as well ( $t(27) = 3.473, p = 0.001$  (1-tailed)). B\_Int paired with B\_Coh showed an almost significant result ( $t(27) = 1.611, p = 0.0595$  (1-tailed)).

The third years' results showed only one significant result in the descriptors of text A. A\_Vocab paired with A\_Int was not significant ( $t(24) = 1.365, p = 0.0925$  (1-tailed)).

A\_Vocab paired with A\_Coh yielded a significant result ( $t(24) = 2.397, p = 0.0125$  (1-tailed)).

A\_Int paired with A\_Coh yielded no significant result ( $t(24) = 1.371, p = 0.0915$  (1-tailed)).

The third years' results for text B did not show any significant results. B\_Vocab paired with B\_Int yielded ( $t(24) = 1.100, p = 0.141$  (1-tailed)). B\_Vocab paired with B\_Coh yielded ( $t(24) = 0.000, p = 0.5$  (1-tailed)). B\_Int paired with B\_Coh yielded ( $t(24) = -1.141, p = 0.1325$  (1-tailed)).

The fifth years' results showed some significant and almost significant results in the descriptors of text A. A\_Vocab paired with A\_Int yielded an almost significant result ( $t(27) = 1.613, p = 0.059$  (1-tailed)). A\_Vocab paired with A\_Coh yielded a significant result ( $t(27) = 2.555, p = 0.085$  (1-tailed)). A\_Int paired with A\_Coh yielded another significant result ( $t(27) = 1.971, p = 0.0295$  (1-tailed)).

The fifth years' results for text B showed some significant results. B\_Vocab paired with B\_Int yielded a significant result ( $t(27) = 1.880, p = 0.0355$  (1-tailed)). B\_Vocab paired with B\_Coh yielded no significant result ( $t(27) = 0.000, p = 0.5$  (1-tailed)). B\_Int paired with B\_Coh yielded a significant result ( $t(27) = -2.353, p = 0.013$  (1-tailed)).

### 4.3 Normality per category and year

**Table 2a. Standard deviations of the results of text A per category and year.**

	<b>Year 1</b>	<b>Year 3</b>	<b>Year 5</b>
<b>Vocab</b>	.962	.881	.992
<b>Int</b>	1.278	.881	.799
<b>Coh</b>	1.274	.913	.997
<b>Overall</b>	1.005	.64	.854

Although the responses were not always the right level, Table 2a shows that the responses are less divided in the third and fifth year than in the first year. This is more clearly visible when year one and year three are compared. A\_Int in year 1 for example shows a standard deviation of 1.278 whilst A\_Int in year three shows a standard deviation of 0.881. The table shows that this same trend can be seen when comparing the data for year one and five. The standard deviation is not always decreasing when comparing year three with year five. This table shows that the results are more normally distributed as the pupils get older, which can be deduced from the standard deviation getting smaller. This means that the distribution of the data seems to centre more on one point in year five and three than in year one.

**Table 2b. Standard deviations of the results of text B per category and year.**

	<b>Year 1</b>	<b>Year 3</b>	<b>Year 5</b>
<b>Vocab</b>	.962	.816	.645
<b>Int</b>	.994	1.036	.92
<b>Coh</b>	1.357	.764	.753

The responses for text B were not always the correct level either, but Table 2b shows that the responses concerning text B seem to centre more on point in year five than in year one. This is clearly visible in B\_Coh where year one shows a standard deviation of 1.357 whilst year three and five show a much smaller standard deviation, 0.764 and 0.753 respectively. Table 2b also shows that a decrease of the standard deviation can be observed when comparing the data of year three and five, which was less clear in Table 2a. Table 2b shows that the results are more normally distributed as the pupils get older, which can be deducted from the standard deviation getting smaller. Interestingly, B\_Int shows a higher standard deviation in year three than in year one. In general however, it can be said that the distribution of the data seems to centre more on one point in year five than in year one.

## 5. Discussion

This chapter will discuss the results that are presented in the previous chapter. The results will be discussed with aid of the hypotheses presented in section 3.5.

### 5.1 The overall levels of text A and B

During the analysis of the data it was noted that the data concerning the overall levels of text A and B lacked significance. This does not mean that the data are unsuitable for this research, but that a different approach was needed in order to be able to deduce any conclusions. It also means that the conclusions drawn are mere assumptions and further research needs to be done in order to come to final conclusions. An interesting finding which was noted whilst conducting the experiments was that the younger the pupils, the less instruction they needed or wanted. The first years were very eager to start and there were hardly any questions during the task, whereas the fifth years required a more extensive instruction and wanted to know more details about how to fill out this task properly. The fifth year pupils also wanted to ask more questions during the task. These questions were often to get a confirmation whether they were filling it out correctly. This may have to do with the pressure to do well, a notion that most first years may not be so familiar with yet. The overall levels of text A and B were not significant, which could be due to many reasons. It is possible that the task itself was too long and that the pupils tended to lose their focus as they progressed to text B. The task took around twenty minutes, which might have been too long, especially for the younger participants. Another possibility might be that the test battery was too small. This research project did not allow for a very large test battery due to limits in time, but a larger amount of participants might be the right way to gain a better overview of the results and their significance. Lastly, it may be possible that the pupils found the texts too difficult, especially text B. Future research will have to run multiple pre-tests in order to find the right texts for a similar experiment. Although a quantitative research appeared to be insufficient in order to find useful results on the overall assigned levels, qualitative research provides some more insight in the data, allowing some conclusions to be drawn from the data.

Table 1 shows that the percentage of pupils who assigned the right overall level was increasing for text A. In year one 17.9% of participants assigned the correct overall level (A2) to the text, whilst in year five 42.9% of the pupils assigned the correct level to the text. This shows that, although this percentage is still lower than predicted, the numbers are increasing and more pupils were able to assign the right level. It is a good thing to see that these results are increasing, but for the CEFR to be used effectively by the pupils, we would have to see

that 80% of year five would have assigned the right level. The percentages for the correct overall level for text B are not at all similar to those of text A. It is observed that fewer pupils assigned the right level as the pupils get older. To clarify, for text B, year one scored 17.9% correct and year five scored 3.8% correct.

These results can be explained by looking at the data more closely. Many pupils seem to overestimate the level of the text in both text A and B. It is possible that pupils take their own level as a reference point and estimate whether the text they have read is better or worse than something they would produce themselves. This implies that a first year pupil estimates that he or she is in an A-level and observes that text B is better than a text that he or she would produce and therefore chooses a higher level, which in this case, is a B-level. This assumption can also be made to clarify the results of the overall level of text B in year five. If the pupil estimates that he or she is at a B-level and estimates that this piece of writing is better than something he or she would produce, it is likely that he or she would choose a C-level. For example, a fifth year pupil who is currently at level B2, might estimate that a text which he thinks is better, should score higher thus C1 at least. Which may not be correct, because the text may not be better at all or only in some categories. A closer look at the data show that many fifth year pupils indeed chose C1 as the correct level. This is just an assumption and in order to come to definite answers more research has to be conducted concerning this issue.

## **5.2 The difficulty of the descriptors**

I deduced from my own experience that pupils would find it easier to assign the right level for vocabulary than to assign the right level for punctuation and coherence, because vocabulary is a topic which is paid a lot of attention to from the first year. This expectation was not borne out, as can be deduced from the data. Vocabulary is perceived far more difficult than punctuation and coherence. The right levels for text A were A2 in all the different categories, and therefore also for the overall level. These data presented more significant results and some almost significant results. This means that these data are less likely to be a coincidence and are therefore able to base conclusions on. Every year scored best for coherence in text A. This means that the mean was closer to the target level for coherence than for any other descriptor. The data show that in text A the pupils were better at assigning the level in coherence than for vocabulary and punctuation, which leads to the assumption that pupils find it easier to evaluate coherence than vocabulary and punctuation. The results for text B lead to a different conclusion. The pupils in year one were again better at assigning the right level in coherence, but the pupils in year three and year five performed better in

punctuation. However, the differences between coherence and punctuation were very small. The mean for coherence in year three was  $M = 4.60$  and for punctuation  $M = 4.36$ , which is a difference of 0.24. In year five the mean for coherence was  $M = 4.75$  and for punctuation  $M = 4.43$ , which is a difference of 0.32. These results show that my initial idea was far from correct. The data show that vocabulary is apparently the most difficult category to evaluate in terms of this experiment. It can be said that following the results of text A, coherence was easier to evaluate followed by punctuation, and the most difficult category to evaluate turned out to be vocabulary. It can then also be said that following the results of text B punctuation was easiest to evaluate followed by coherence and the most difficult to evaluate turned out to be vocabulary. This might have to do with the way pupils look at a text. It is possible that pupils look at a text more as a whole or as a piece of text on a paper, rather than what the text is actually about and how it is written. Another possibility is that the pupils do not have a wide range of vocabulary themselves, or that the descriptors in this category are more vaguely described than the descriptors concerning coherence and punctuation. In order to find out what the reason is for vocabulary being the most difficult category to judge, further research in this particular subject needs to be conducted.

A qualitative analysis led to an interesting observation in the data of text B in year three and five. The data show that vocabulary and coherence in the B text have the same mean of 4.60 in year three and 4.75 in year five. These results are far from the target result, which was set for text B at 3 (B1), but it does show that apparently pupils in year three and five found it as difficult to evaluate coherence as to evaluate vocabulary. What further implications this has is a topic for further investigation, but it might mean that pupils connect coherence with vocabulary as they grow older, or that vocabulary and coherence start to be more difficult to evaluate or are more closely related to each other as the level of the text increases.

### **5.3 Normality per category and year**

Table 2a shows the standard deviations regarding the given answers for each individual category as well as the overall level per year for text A. This table shows a development in the way the pupils evaluate the text and although the distribution of the results were not included in the original hypotheses, it is too interesting not to pay any attention to. What can be observed in this table is that the first years are giving answers of which the data have a large standard deviation, meaning that answers are more evenly divided over all the possible answer options. This implies that first year pupils do not have a very well developed judging skill when looking at the entire year. This is not very surprising since they have not

often encountered the CEFR, nor are they very skilled in reading and writing English. Third year pupils start to show a smaller standard deviation. This means that pupils in their third year have a better understanding of how to evaluate a piece of writing and have had more training from previous years with regards to reading and writing English. The fifth year pupils show even more normally distributed results, which probably has to do with their proficiency in writing and reading English and their ability to evaluate a piece of writing. The fifth year pupils are also more likely to know more about the CEFR and to have used the CEFR more often than first year pupils.

The same thing can be said about table 2b. The older the pupils are, the more they seem to unify in their results. The fifth year pupils show smaller standard deviations than the first year pupils. B\_Coh, for example, year one shows a standard deviation of 1.357 whilst year three and five show a much smaller standard deviation, 0.764 and 0.753 respectively. This shows that many pupils chose different levels. Therefore, there is no unification in the results. In general, it can be said that the distribution of the data seems to centre more on one point in year five than in year one in the results of text A as well as in the results of text B.

It is important to note that this is not about choosing the correct level, but about the distribution of the results. When there is a normal distribution in results, it means that people start to think more alike and thus the results seem to centre more on one point. Fifth year pupils have had many more English classes and are more trained in looking critically at a piece of work than first year pupils. It seems therefore that length of education in years has an effect on how much pupils start to or are trained to think as their peers and that education leads to unification. However, to prove this point, further research has to be conducted in the field of sociology or psychology with this particular relation in mind, for which it is not important what an individual answers, but what the group as a whole answers.

## 6. Conclusion

This thesis was built upon the research question whether second language learners can use the CEFR effectively as a tool for evaluation. In order to find an answer to the main hypothesis, sub-hypotheses were constructed to help analyse the data and draw conclusions that either support or oppose the main hypothesis. These hypotheses were tested through an experiment which asked pupils in years one, three, and five to evaluate two pieces of writing according to the CEFR scales in three different categories. These three categories were vocabulary, punctuation, and coherence. The levels that needed to be assigned in order to be correct were A2 for text A en B1 for text B. These results were collected and underwent a quantitative analysis using *IBM SPSS* and a qualitative analysis in order to get some more insight in the results. All of these data combined lead to the following conclusions.

The first set of sub-hypotheses sought to find an answer to the question whether pupils in years one, three, and five were capable of assigning the correct overall level in text A and B. The reason for giving the pupils two texts to evaluate is to ensure that every group of pupils had a text that was a bit of a challenge for them. However, the quantitative analysis did not yield any significant results, but a qualitative analysis of the data showed interesting results. These results showed that many pupils had difficulties with assigning the correct overall level, but that the number of pupils who are able to assign the right overall level increased as the pupils get older. In year one only 17.9% was able to assign the right level for text A, whereas in year five 42.9% was able to do so. This means that the pupils do get better in using the CEFR as a tool for evaluation over time. This may have to do with a higher proficiency in reading and writing English, as well as a higher ability to look critically at a piece of writing. The extra years of education have increased their level of English and helped them to critically review texts. Another result that became clear after the qualitative analysis was that the pupils systematically overestimated the pieces of writing. Text B was supposed to be assigned at a B1 level, but most pupils chose for a C1 or C2 level. This over estimation of the level is a clear indication that pupils struggle to assign the right level and it also shows that a lot of pupils think very highly of a text they read.

The second set of sub-hypotheses sought to find an answer to the question whether there was a specific category that pupils found difficult to evaluate. The expectation was that pupils would find it easier to evaluate vocabulary rather than punctuation or coherence. The data that were analysed showed some interesting and significant results regarding this topic. Pupils performed far better when assigning the right level for coherence in text A. This was surprising, because the expectation was that coherence would be one of the most difficult

categories. Text B showed that punctuation was the easiest category to evaluate correctly. It turned out that vocabulary was, in all cases, the most difficult category to evaluate. Coherence and punctuation being easier categories to evaluate might have to do with the way pupils look at a text. Pupils may look at a text more as a whole, rather than what the text is actually about and how it is written. Evaluating something based on visual cues may be easier for pupils than an evaluation based on the contents of a text, but this has to be researched further in order to draw any conclusions regarding this matter.

The last results that were found were not described in any sub-hypothesis, but these shed an interesting light on the evaluation skills of pupils. As the pupils get older, the pupils do not only seem to get better at assigning the right level, the standard deviations of the results also become smaller in most cases, meaning that the answers are less spread out over all the possibilities and students answer more uniformly. The first year pupils assigned levels in a range of all the six different CEFR levels, whereas the fifth year pupils assigned levels within four of the possible levels. It is important to note that this is not about assigning the correct level, but about the distribution of the results. When the standard deviations are smaller, it means that people start to think more alike and thus the results seem to centre more on one point. Fifth year pupils have had many more English classes and are more trained in looking critically at a piece of work than first year pupils. The idea is that when a pupil has had more training in a certain language in several fields, it may get easier to more thoroughly understand a text which is needed to evaluate a text correctly. Although the assigned were not always correct, the tables show that the answers seem to centre more on one point. This could be a sign that the pupils are getting better in evaluating texts.

Bearing these conclusions to the sub-hypotheses in mind, it is possible to draw a conclusion concerning the main hypothesis of this thesis. The results show that pupils find it difficult to use the CEFR effectively as a tool for evaluation. There were not many pupils in the first year who assigned the right level and although the number increased as the pupils get older, the number of pupils who assigned the right level were still quite low. Using the CEFR as a tool for evaluation appeared to be more difficult when looking at vocabulary. Coherence and punctuation on the other hand were considered slightly easier to assign in this experiment. It also showed that pupils start to answer more uniformly as they get older.

In order to make CEFR a more useful tool for evaluation for pupils, it might be necessary to build the descriptors in a way that is more transparent for pupils. On the one hand, this could be done by adding examples to the descriptors to make them more clear. On the other hand, teachers can play a role in this as well by taking up courses in the curriculum

which teach the pupils how to use the CEFR more successfully in several areas such as evaluation. Pupils being able to evaluate themselves successfully according the CEFR model can be helpful to pupils, because it enables to see on which level they currently are and what they need to do in order to improve themselves.

### **6.1 Flaws and future research**

In hindsight, the experiment for this research had some flaws and can be improved in several areas. The test battery for this experiment was not very large due to scope of this thesis. This in turn contributed to the fact that the results were not always significant. This is a part that can easily be improved by gathering a larger amount of participants. It may also be that the texts used in this experiment were not ideal. It could be that the texts were too difficult or that the pupils would have benefitted from having the original writing prompts as well. The experiment may also be conducted with descriptors from different categories. It could be that these categories were ill-chosen and that different categories yield better results. It might therefore be useful to design several experiments with different texts, different categories and with or without the writing prompts and to run pilot experiments on small groups to get an indication of what will probably yield the most useful results.

This research, however, has opened up some new insights and raised some more issues which can be looked at or investigated in the future. The results have shown that pupils do get better at assigning the right level, but it is unclear whether this can be said for the whole population of the Netherlands or for the European population for that matter. It may also be interesting to also have a look at the different levels in secondary school rather than only looking at VWO. This may provide some insights in what pupils from different levels find difficult.

Another subject that future research may want to explore is the different categories and which category is thought to be the most difficult and why. Figuring out exactly where the difficulties are may be the next step in tackling these difficulties and that can be a big step forward in making the CEFR a more effective tool for pupils.

An international study could be set up in order to find how big the influence is on how the pupils are taught to evaluate a piece of writing. The data from this research showed that the results of year five seemed to centre more on one point, but were still often incorrect. It might be the case that other countries have other strategies which are beneficial for the pupils in terms of being more capable of evaluating work.

Lastly, this research assumed that if pupils can look at written production by someone

they do not know and evaluate them correctly according to the CEFR, they can evaluate themselves too. This is a rather large assumption and further research has to be done in order to find whether this assumption is at all true or not. It is possible that there is a difference between the different ages in secondary schools. First year pupils may not be able to look at their own work critically, because they lack a wide knowledge on the topic or lack a higher proficiency of the language they are writing in. An issue which becomes smaller over the years. Another reason why first year pupils may not be able to look critically at their own work could be found in the psychological development of children. It may be possible that self-reflection starts to develop significantly from a certain age. This assumption is an issue that can be researched in several areas and might even differ per school or per language. It is also possible that this assumption was correct, and if it is not, it might provide some more insights in what pupils find difficult in evaluating written work in foreign languages. Although some of the results were not significant, this thesis shed some new light on issues concerning the CEFR and provides new explorative topics for future research.

## 7. References

- Alderson, J. C. (2007). The CEFR and the Need for More Research. *The Modern Language Journal*, 91(4), 659–663. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626093>
- Bonnet, G. (2007). The CEFR and Education Policies in Europe. *The Modern Language Journal*, 91(4), 669–672. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626096>
- Europees Referentiekader Talen. (no date). *Engels Schrijven A2*. Consulted on 8 May 2016, retrieved from <http://www.erk.nl/leerling/Voorbeelden/Engels/en-sch-01/>
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics: and Sex and Drugs and Rock 'n' Roll*. London: Sage.
- Figueras, N. (2007). The CEFR, a Lever for the Improvement of Language Professionals in Europe. *The Modern Language Journal*, 91(4), 673–675. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626097>
- Hulstijn, J. H. (2007). The Shaky Ground beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency. *The Modern Language Journal*, 91(4), 663–667. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626094>
- Krumm, H. J. (2007). Profiles Instead of Levels: The CEFR and Its (Ab)Uses in the Context of Migration. *The Modern Language Journal*, 91(4), 667–669. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626095>
- Larsen-Freeman, D., & Freeman, D. (2008). Language Moves: The Place of "Foreign" Languages in Classroom Teaching and Learning. *Review of Research in Education*, 32, 147–186. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/20185115>
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Language Education Policy. *The Modern*

- Language Journal*, 91(4), 645–655. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626091>
- Mislevy, R. J. (1995). Test Theory and Language Learning Assessment. *Language Testing* 12, 3, 341–369.
- Moonen, M., Stoutjesdijk, E., Graaff, R. D., & Corda, A. (2010). Het ERK in het Voortgezet Onderwijs: Ervaringen van Docenten Moderne Vreemde Talen. *Levende Talen Tijdschrift*, 11(4), 34-44. Retrieved from <http://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/89>
- North, B. (2007a). The CEFR Illustrative Descriptor Scales. *The Modern Language Journal*, 91(4), 656–659. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626092>
- North, B. (2007b). The CEFR: Development, Theoretical and Practical Issues. *Babylonia*, 1, 22-29. Retrieved from [http://babylonia.ch/fileadmin/user\\_upload/documents/2007-1/Baby2007\\_1North.pdf](http://babylonia.ch/fileadmin/user_upload/documents/2007-1/Baby2007_1North.pdf)
- North, B. (2014). *The CEFR in Practice* (Vol. 4). Cambridge University Press
- Stevens, D. (2015). *A Teacher's Profession Is an Ideal Profession. Yet All Teachers Are Not Ideal*. Assignment. Database Dr. C.M. de Vries.
- Trim, J. L. (2012). The Common European Framework of Reference for Languages and its Background: A Case Study of Cultural Politics and Educational Influences. *The common European framework of reference: The globalisation of language education policy*, 14-34.
- Westhoff, G. (2007). Challenges and Opportunities of the CEFR for Reimagining Foreign Language Pedagogy. *The Modern Language Journal*, 91(4), 676–679. Retrieved from <http://www.jstor.org.ru.idm.oclc.org/stable/4626098>

## 8. Appendices

### A - Experiment text A

#### How Many Ways Do I Love Thee?

*Every year in the beginning of February, every shop has cute red stuff in their showroom. But what want people in love to buy? I have done a searching about people's wishes for Valentine's Day. I asked many people what they want to buy for their valentine.*

*I got the following answers.*

*38% of the people, which I asked about their purchases, have bought candy. That holds red sweet hearts, red lollypops and little pink hearts. 44% chose a Date Night what means a night in a beautiful romantic hotel.*

*32% buys flowers for their lover. Red flowers, red roses or white roses.*

*29% gives there valentine gift cards, for example a gift card for a book or a card for a day in a health enter.*

*They chose the worth of the card what they want to give.*

*65% of the people send a post card. They will tell in this way their love for the receiver.*

*A little percent, 11 %, buys jewellery for their love. In the most cases those jewellery are very expensive.*

*17 % are very creative and buys other things than I have shown on the graph.*

*12 % spend a lot of money on expensive perfumes and colognes.*

*17 % gives their love cute plush.*

*That is the information that my graph shows.*

*The horizontal axe shows the households participation rates and the staves show how much the people that I have asked have bought.*

**B - Experiment text B**

*A Teacher's Profession Is an Ideal Profession. Yet All Teachers Are Not Ideal.'*

Teachers are important in the life of children. Parents raise their children and teachers educate them. This is true but, teachers do more than educate children. At least, a good teacher does. Children, especially young ones are majorly influenced by their teachers. The way the teacher interacts with them and how they are educated affects their life in a big way. But as quoted above, what makes a teacher an ideal one? In my opinion teaches a good teacher students that learning is fun, pushes them to want to do their best and loves what he or she does.

One thing a good teacher does is showing students that learning can be fun. As soon as someone has fun with learning, the process gets way more effective. People who enjoy learning usually tend to learn faster too. This is why teachers should always at least try to show their students that learning can be fun.

A good teacher also pushes their students to do the best they can do. Challenging tends to get the best results out of students, out of everyone even. Not being challenged tends to students getting bored and that usually results in underachievement. By being challenged most people will strive to achieve the best result, and a teacher challenging students will get the best results.

Someone who is not passionate about something will not get other people excited for it , while if they are passionate, they will easily get you excited for it too. This is the same for teachers, if a teacher is not passionate about what they are teach, students will also not be motivated for the subject. Passionate teachers will pass their passion on to their students.

Of course it is impossible to say 'if you do this you will be a perfect teacher' because it depends on the group of students. Everyone has a different opinion on what makes a 'good' teacher and what makes a 'bad' teacher. And in every way the 'ideal' teacher is a different combination of characteristics. Still, the key element between a good teacher and a teacher is showing students that learning does not mean memorizing texts form a book and challenge them to strive for the best result. Being passionate about your subject and passing this on to students and supporting them is also very important. *'A teacher is never the giver of truth; he is a guide, a pointer to the truth each student must find for himself'* as Bruce Lee once said.

**C - Complete experiment****Leeftijd:****Klas:****Geslacht M / V****Wat moet je doen?**

1. Lees wat de leerling heeft ingeleverd.
2. Kruis aan op welk niveau jij vindt dat de leerling zit. Per gebied (Woordenschat, spelling en samenhang) zet je een kruis in welk niveau deze leerling past volgens jou.
3. Geef een toelichting of beargumentatie voor je keuze. (voorbeelden uit de tekst).

Zie ook onderstaand voorbeeld:

Als je bijvoorbeeld kiest voor niveau B1 bij woordenschat, ziet je ingevulde schema er zo uit:

<b>BEREIK VAN DE WOORDENSCHAT</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>B1</b>	Beschikt over een voldoende woordenschat om zich, met enige omhaal van woorden, te uiten over de meeste onderwerpen die betrekking hebben op het dagelijks leven, zoals familie, vrijetijdsbesteding en interesses, werk, reizen en actualiteiten	x	two brothers, a sister and a dog.  I like snowboarding and horse riding like reading and shopping.  Spain in the summer holidays.

Beoordeel het niveau van deze prestatie op de volgende pagina's.

### **How Many Ways Do I Love Thee?**

*Every year in the beginning of February, every shop has cute red stuff in their showroom. But what want people in love to buy? I have done a searching about people's wishes for Valentine's Day. I asked many people what they want to buy for their valentine.*

*I got the following answers.*

*38% of the people, which I asked about their purchases, have bought candy. That holds red sweet hearts, red lollypops and little pink hearts. 44% chose a Date Night what means a night in a beautiful romantic hotel.*

*32% buys flowers for their lover. Red flowers, red roses or white roses.*

*29% gives there valentine gift cards, for example a gift card for a book or a card for a day in a health enter.*

*They chose the worth of the card what they want to give.*

*65% of the people send a post card. They will tell in this way their love for the receiver.*

*A little percent, 11 %, buys jewellery for their love. In the most cases those jewellery are very expensive.*

*17 % are very creative and buys other things than I have shown on the graph.*

*12 % spend a lot of money on expensive perfumes and colognes.*

*17 % gives their love cute plush.*

*That is the information that my graph shows.*

*The horizontal axe shows the households participation rates and the staves show how much the people that I have asked have bought.*

**Beoordelingsformulier: How Many Ways Do I Love Thee?**

<b>BEREIK VAN DE WOORDENSCHAT</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Heeft een zeer elementaire woordenschat die bestaat uit geïsoleerde woorden en eenvoudige uitdrukkingen met betrekking tot persoonlijke gegevens en bepaalde concrete situaties.		
<b>A2</b>	Beschikt over voldoende woordenschat om zich te redden bij primaire levensbehoeften.		
<b>B1</b>	Beschikt over een voldoende woordenschat om zich, met enige omhaal van woorden, te uiten over de meeste onderwerpen die betrekking hebben op het dagelijks leven, zoals familie, vrijetijdsbesteding en interesses, werk, reizen en actualiteiten		
<b>B2</b>	Beschikt over een voldoende brede woordenschat voor zaken die verband houden met zijn of haar vakgebied en de meeste algemene onderwerpen. Kan variatie aanbrengen in formuleringen om te veel herhaling te voorkomen, al		

	kunnen tekortkomingen in de woordenschat nog wel tot omschrijving leiden		
<b>C1</b>	Heeft een goede beheersing van een breed repertoire aan woorden, waardoor tekortkomingen in de woordenschat gemakkelijk kunnen worden gedicht met omschrijvingen; er is in geringe mate sprake van zichtbaar zoeken naar uitdrukkingen. Heeft een goede beheersing van idiomatische uitdrukkingen en uitdrukkingen uit de spreektaal		
<b>C2</b>	Heeft een goede beheersing van een heel breed lexicaal (woordenschat) repertoire met inbegrip van idiomatische uitdrukkingen en uitdrukkingen uit de spreektaal; toont zich bewust van verschil in betekenissen		

<b>SPELLING, INTERPUNCTIE EN LAY-OUT</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Kan vertrouwde woorden en korte zinsdelen overschrijven, bijvoorbeeld eenvoudige aanwijzingen of instructies, namen van alledaagse dingen, namen van winkels en regelmatig gebruikte standaardzinnnetjes . Kan zijn of haar adres, nationaliteit, en andere persoonlijke gegevens spellen		
<b>A2</b>	Kan korte zinnen over alledaagse onderwerpen overschrijven, bijvoorbeeld een routebeschrijving. Kan fonetisch (zoals het klinkt) redelijk correct (maar niet noodzakelijkerwijs helemaal in de standaardspelling) korte woorden opschrijven uit zijn of haar gesproken woordenschat		
<b>B1</b>	Kan heldere doorlopende tekst produceren die over het algemeen helemaal te begrijpen is. Spelling, leestekengebruik en lay-out zijn correct genoeg om het grootste deel van de tijd te kunnen worden gevolgd		

<b>B2</b>	Kan heldere, begrijpelijke, doorlopende tekst produceren die voldoet aan standaardconventies voor lay-out en alinea indeling. Spelling en leestekengebruik zijn redelijk correct maar kunnen invloeden van moedertaal verraden		
<b>C1</b>	Lay-out, alinea indeling en leestekengebruik zijn consistent en bevorderen de leesbaarheid. De spelling is correct, afgezien van een enkele verschrijving		
<b>C2</b>	Het geschrevene is qua spelling, lay-out en alinea indeling foutloos		

### SAMENHANG

<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Kan woorden of woordgroepen met elkaar verbinden met heel elementaire verbindingswoorden als 'en' of 'dan' (and, then/than)		
<b>A2</b>	Kan woordgroepen verbinden met eenvoudige verbindingswoorden als 'en', 'maar' en 'omdat'. (and, but, because)		

<b>B1</b>	Kan een reeks kortere, op zichzelf staande eenvoudige elementen verbinden tot een samenhangende opeenvolging van punten (argumenten)		
<b>B2</b>	Kan gebruikmaken van een beperkt aantal samenhang bevorderende elementen om zijn of haar uitingen te verbinden tot een helder, samenhangend betoog, hoewel in een langere bijdrage sprake kan zijn van een zekere 'springerigheid'.		
<b>C1</b>	Kan duidelijke, vloeiende, goed gestructureerde gesproken taal voortbrengen, die getuigt van een goede beheersing van ordeningspatronen, verbindingswoorden en samenhang		
<b>C2</b>	Kan samenhangende tekst produceren en daarbij volledig en gepast gebruik maken van uiteenlopende ordeningspatronen en een breed scala van samenhang bevorderende elementen		

Beoordeel het niveau van deze prestatie op de volgende pagina's.

*A Teacher's Profession Is an Ideal Profession. Yet All Teachers Are Not Ideal.'*

Teachers are important in the life of children. Parents raise their children and teachers educate them. This is true but, teachers do more than educate children. At least, a good teacher does. Children, especially young ones are majorly influenced by their teachers. The way the teacher interacts with them and how they are educated affects their life in a big way. But as quoted above, what makes a teacher an ideal one? In my opinion teaches a good teacher students that learning is fun, pushes them to want to do their best and loves what he or she does.

One thing a good teacher does is showing students that learning can be fun. As soon as someone has fun with learning, the process gets way more effective. People who enjoy learning usually tend to learn faster too. This is why teachers should always at least try to show their students that learning can be fun.

A good teacher also pushes their students to do the best they can do. Challenging tends to get the best results out of students, out of everyone even. Not being challenged tends to students getting bored and that usually results in underachievement. By being challenged most people will strive to achieve the best result, and a teacher challenging students will get the best results.

Someone who is not passionate about something will not get other people excited for it , while if they are passionate, they will easily get you excited for it too. This is the same for teachers, if a teacher is not passionate about what they are teach, students will also not be motivated for the subject. Passionate teachers will pass their passion on to their students.

Of course it is impossible to say 'if you do this you will be a perfect teacher' because it depends on the group of students. Everyone has a different opinion on what makes a 'good' teacher and what makes a 'bad' teacher. And in every way the 'ideal' teacher is a different combination of characteristics. Still, the key element between a good teacher and a teacher is showing students that learning does not mean memorizing texts form a book and challenge them to strive for the best result. Being passionate about your subject and passing this on to students and supporting them is also very important. *'A teacher is never the giver of truth; he is a guide, a pointer to the truth each student must find for himself'* as Bruce Lee once said.

**Beoordelingsformulier: A Teacher's Profession**

<b>BEREIK VAN DE WOORDENSCHAT</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Heeft een zeer elementaire woordenschat die bestaat uit geïsoleerde woorden en eenvoudige uitdrukkingen met betrekking tot persoonlijke gegevens en bepaalde concrete situaties.		
<b>A2</b>	Beschikt over voldoende woordenschat om zich te redden bij primaire levensbehoeften.		
<b>B1</b>	Beschikt over een voldoende woordenschat om zich, met enige omhaal van woorden, te uiten over de meeste onderwerpen die betrekking hebben op het dagelijks leven, zoals familie, vrijetijdsbesteding en interesses, werk, reizen en actualiteiten		
<b>B2</b>	Beschikt over een voldoende brede woordenschat voor zaken die verband houden met zijn of haar vakgebied en de meeste algemene onderwerpen. Kan variatie aanbrengen in formuleringen om te veel herhaling te voorkomen, al		

	kunnen tekortkomingen in de woordenschat nog wel tot omschrijving leiden		
<b>C1</b>	Heeft een goede beheersing van een breed repertoire aan woorden, waardoor tekortkomingen in de woordenschat gemakkelijk kunnen worden gedicht met omschrijvingen; er is in geringe mate sprake van zichtbaar zoeken naar uitdrukkingen. Heeft een goede beheersing van idiomatische uitdrukkingen en uitdrukkingen uit de spreektaal		
<b>C2</b>	Heeft een goede beheersing van een heel breed lexicaal (woordenschat) repertoire met inbegrip van idiomatische uitdrukkingen en uitdrukkingen uit de spreektaal; toont zich bewust van betekenisconnotaties		

<b>SPELLING, INTERPUNCTIE EN LAY-OUT</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Kan vertrouwde woorden en korte zinsdelen overschrijven, bijvoorbeeld eenvoudige aanwijzingen of instructies, namen van alledaagse dingen, namen van winkels en regelmatig gebruikte standaardzinnnetjes . Kan zijn of haar adres, nationaliteit, en andere persoonlijke gegevens spellen		
<b>A2</b>	Kan korte zinnen over alledaagse onderwerpen overschrijven, bijvoorbeeld een routebeschrijving. Kan fonetisch (zoals het klinkt) redelijk correct (maar niet noodzakelijkerwijs helemaal in de standaardspelling) korte woorden opschrijven uit zijn of haar gesproken woordenschat		
<b>B1</b>	Kan heldere doorlopende tekst produceren die over het algemeen helemaal te begrijpen is. Spelling, leestekengebruik en lay-out zijn correct genoeg om het grootste deel van de tijd te kunnen worden gevolgd		

<b>B2</b>	Kan heldere, begrijpelijke, doorlopende tekst produceren die voldoet aan standaardconventies voor lay-out en alinea indeling. Spelling en leestekengebruik zijn redelijk correct maar kunnen invloeden van moedertaal verraden		
<b>C1</b>	Lay-out, alinea indeling en leestekengebruik zijn consistent en bevorderen de leesbaarheid. De spelling is correct, afgezien van een enkele verschrijving		
<b>C2</b>	Het geschrevene is qua spelling, lay-out en alinea indeling foutloos		

<b>SAMENHANG</b>			
<b>Niveau</b>	<b>Beschrijving</b>	<b>x</b>	<b>Voorbeelden uit de tekst</b>
<b>A1</b>	Kan woorden of woordgroepen met elkaar verbinden met heel elementaire verbindingswoorden als 'en' of 'dan' (and, then/than)		
<b>A2</b>	Kan woordgroepen verbinden met eenvoudige verbindingswoorden als 'en', 'maar' en 'omdat'. (and, but, because)		

<b>B1</b>	Kan een reeks kortere, op zichzelf staande eenvoudige elementen verbinden tot een samenhangende opeenvolging van punten		
<b>B2</b>	Kan gebruikmaken van een beperkt aantal samenhang bevorderende elementen om zijn of haar uitingen te verbinden tot een helder, samenhangend betoog, hoewel in een langere bijdrage sprake kan zijn van een zekere 'springerigheid'.		
<b>C1</b>	Kan duidelijke, vloeiende, goed gestructureerde gesproken taal voortbrengen, die getuigt van een goede beheersing van ordeningspatronen, verbindingswoorden en samenhang		
<b>C2</b>	Kan samenhangende tekst produceren en daarbij volledig en gepast gebruik maken van uiteenlopende ordeningspatronen en een breed scala van samenhang bevorderende elementen		