

The Correlation between Self-Assessment Reports and CEFR Levels according to Standardized Assessment

The DIALANG Vocabulary Test and the Cambridge General English Online Test in relation to two Self-Assessment Procedures: Likert Scale Ratings and the Employment of CEFR Descriptors

Master's Thesis – MA Linguistics: Language and Communication Coaching
Radboud University, Nijmegen

Student: Laura Buuts

Student Number: 4423194

Date of Submission: 21-06-2020

Supervisor: J. Klatter-Folmer

Second Reader: E. Krikhaar

Radboud University



ABSTRACT

This thesis concerns an investigation into the correlation between self-assessment reports regarding vocabulary knowledge and actual levels of vocabulary knowledge according to two standardized assessments. The experiment in this study was conducted by means of a questionnaire consisting of a language testing component and a self-assessment component. The participants in this study are speakers of Dutch as a native language. With the help of regression analyses, it has been shown that the result of this study is that the Common European Framework of Reference (CEFR) levels of vocabulary knowledge according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test correlate with the estimated levels of the self-assessment component. There are differences in the correlation coefficients of both tests and the self-assessment components. Unfortunately, the sample size impeded the confirmation of demographic discrepancies, such as gender, educational background, and use of English in daily life and in the workplace or educational context. Nonetheless, the fact that in general the language testing component correlates with the self-assessment component indicates that in terms of vocabulary, participants are generally able to predict their language skills. This could have some implications for self-assessment as an assessment tool in educational contexts, or for other contexts where language testing is applicable.

Table of Content

1. Introduction	1
2. Theoretical Frame	2
<i>2.1. Defining and Specifying L2 Language Proficiency</i>	2
2.1.1 Cummins' 'Iceberg' Representation of L1 and L2 Proficiency	3
2.1.2 Krashen's Monitor Model	5
2.1.3 Vocabulary Knowledge according to CEFR Language Standards	7
<i>2.2 The Correlation between Standardized Assessment and Self-Assessment</i>	10
2.2.1 Assessing Vocabulary Knowledge with Standardized Assessment	11
2.2.2 Delineating the Concept of Self-Assessment	13
2.2.3 Self-Assessment and its Affiliation to Standardized Assessment	16
3. Research Questions and Hypotheses Development	20
3.1 Research Questions	20
3.2 Hypotheses Development	20
4. Methodology	23
4.1 Participants	24
4.2 Instrumentation	25
4.3 Procedure	28
4.4 Design	29
4.5 Qualitative and Statistical Analyses	29
5. Results	30

<i>5.1 Distribution of the Selected Answer Options in the Language Testing Component and the Self-Assessment Component</i>	31
5.1.1 Report of the Provided Answers to the DIALANG Vocabulary Test	31
5.1.2 Report of the Provided Answers to the Cambridge General English Online Test	46
<i>5.2 Correlations between the post-test Self-Assessment Component and the CEFR Levels of Vocabulary Knowledge</i>	53
5.2.1 Correlations between the post-test Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test	53
5.2.2 Correlations between the post-test Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test	61
<i>5.3 Demographic Discrepancies regarding the Correlations between the Self-Assessment Component and the Language Testing Component</i>	69
5.3.1 Gender	69
5.3.2 Educational Background	73
5.3.3 Use of English	76
6. Discussion	84
6.1 Theoretical and Practical Implications	84
6.2 Future research	87
7. Conclusion	87
8. References	89
9. Appendix	94
9.1 Questionnaire: Engelse woordenschatkennis	94
9.2 Answer Grid: Dialang Vocabulary Test and Cambridge General English Online Test	106
9.3 The Dutch Educational System	107

1. Introduction

The current thesis will examine whether the results on vocabulary tests correlate with particular self-assessment reports. The concept of self-assessment is often used in the educational context, for instance as a tool for students to be placed in a class at a certain level. What is more, self-assessment enables students to measure their current level of competence in different skills, and compare it with their starting level and their target level (Blue, 1994: 5). For students enrolled in a class, this procedure can help them evaluate the progress they have made. Since self-assessment is carried out through complex cognitive processes that are affected by uncontrollable factors, it could not be stated with certainty that the use of self-assessment is effective (Saito, 2003: para 10). Hence, this could result in an inflation bias where one tends to presents one's self in the best light (Delgado, Guerrero, Goggin & Ellis, 1999: 32). However, the literature shows that, overall, language learners tend to report low self-estimates in pronunciation and grammar, and high self-estimates in communicative skills (Blanche, 1988: 82).

There are several studies in the language testing research field that have quantitatively compared self-assessment and objective measures of language proficiency. Although most of these studies demonstrate that self-assessment “[...] tends to carry about the same weight as any of the various parts (sub-tests) of a standardized testing instrument” (Blanche, 1988: 81), more elaborate statistical analyses show that the accuracy of self-assessment could not always be accounted for. This could for instance be due to subjective orders, such as “[...] past academic record, career aspirations, peer-group or parental expectations, lack of training in self-study and self-management” (Blanche, 1988: 81). The author of the latter article advices to employ self-assessment instruments that “[...] contain descriptions of concrete linguistic situations which the learner can size up in behavioral terms” (Blanche, 1988: 82) in order to yield the most accurate self-assessment results. A very important finding for the course of the current thesis is that “[...] higher correlations were obtained between self-assessments based on such situational models and other examination results than between other examination results and global self-appraisals of “macro skills” like “writing”, or “understanding a native speaker” (Blanche, 1988: 81-82). Therefore, in the present thesis, it has been chosen to employ both descriptions of concrete linguistic situations (i.e. the CEFR descriptors) and global self-

appraisals (i.e. the Likert scale ratings). The present study quantitatively compares the results on the language testing component of the employed questionnaire with the results on the self-assessment component of the questionnaire. It was chosen to adopt both the DIALANG Vocabulary Test and the Cambridge General English Online Test in the language testing component. For the self-assessment component, the self-assessment items consist of CEFR descriptors of vocabulary knowledge, and Likert scale ratings of two productive skills (writing ability and oral proficiency), and two receptive skills (listening comprehension and reading ability). The questionnaire was distributed to the general public. It is aimed to evince that the language testing component correlates with the self-assessment component, in order to claim that the subjects that partook in the study are able to accurately report on their language skills after having partaken in a language test. The relevance of this study lies in the implications the correlations have for both tests, and for both types of self-assessment items with regard to language assessment.

In the next chapter, a theoretical frame will be composed for building the theoretical implications that result in the research questions and the hypotheses in chapter 3. In the fourth chapter, the methodology will be outlined so that the statistical analyses could be performed in the fifth chapter. Chapter 6 will report on these statistical analyses in order to strike a balance and review whether the results are in fact significant, and the hypotheses should be accepted or rejected. The last chapter will summarize the proceedings of the present thesis, and the conclusions that could be drawn from the results.

2. Theoretical Frame

This theoretical frame is divided into two main sections. The first section will elaborate on some key concepts with regard to second language (L2) proficiency: Cummins' 'iceberg' model; Krashen's Monitor model, and language proficiency according to the Common European Framework of Reference (CEFR). The second section will be concerned with both standardized assessment and self-assessment, and the possible affiliation between these two types of assessment.

2.1. Defining and Specifying L2 Language Proficiency

This section consists of three sub-sections. The first section will delve into aspects about L1 and L2 proficiency regarding cognitive processes, as proposed by Cummins' 'iceberg' model,

the second section will elaborate on Krashen's Monitor model, and the third section will elaborate on some core concepts with regard to charting vocabulary knowledge according to the Common European Framework of Reference (CEFR).

2.1.1 Cummins' 'Iceberg' Representation of L1 and L2 Proficiency

Before moving on to how the assessment of language proficiency comes along, it should be clarified how language proficiency of the first language (L1) and an additional language are cognitively situated. Regarding the development of language proficiency, Oller (1978) reports on the existence of "[...] a global language proficiency factor which accounts for the bulk of the reliable variance in a wide variety of language proficiency measures" (Oller, 1978: 413). This global language proficiency factor relates to cognitive capacity and aspects of academic achievement. According to Cummins (1980), a large body of research supports this finding in the sense that there seem to exist "[...] high correlations between literacy skills and general intellectual skills" (Cummins, 1980: 84). However, as Cummins (1980) postulates, not the entire scope of language proficiency correlates with cognitive skills, e.g. in the case of mental limitations. In this sense, there seem to be basic interpersonal communicative skills (BICS) in an L1, regardless of mental capacity. While Oller (1978) refers to a 'global language proficiency', Cummins (1978) differentiates BICS from 'cognitive/academic language proficiency' (CALP) which refers to the aspect of language proficiency that is closely linked to literacy skills. In order to visualize the distinction between CALP and BICS, Cummins (1980) has adapted the 'iceberg model' (see Figure 1), originally proposed by Shuy (1976). This model illustrates that grammar, vocabulary, and pronunciation (BICS) are a visible aspect of language proficiency, i.e. above the surface, and that CALP is a dimension of language proficiency that is situated below the surface, which implies the manipulation of language in a decontextualized academic context (Cummins, 1980: 84).

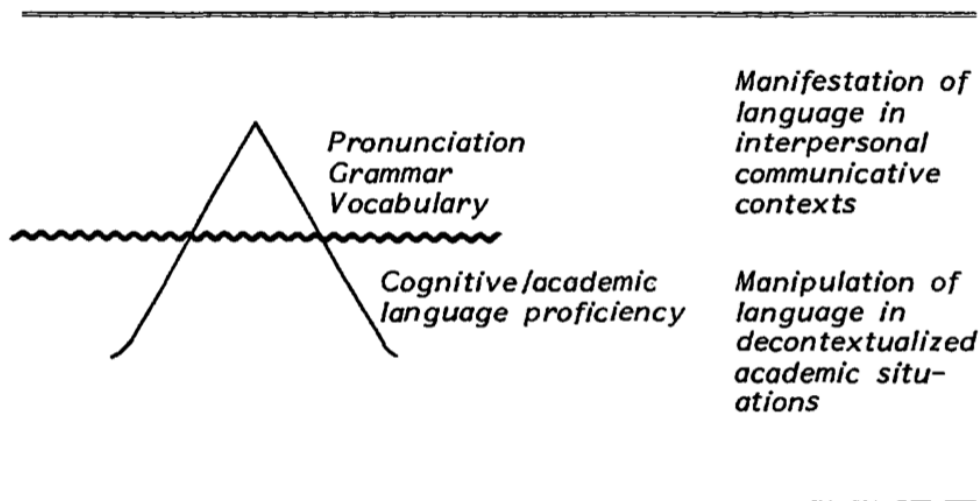


Figure 1: The 'Iceberg' Representation of Language Proficiency (see Cummins, 1980: 84).

At that time of writing, Cummins (1980) stated that there was relatively little research into what forms of language proficiency are associated with the development of literacy skills in school contexts, and how the development of academic skills in L1 is related to the development of academic skills in L2 (Cummins, 1980: 83). All the same, it is to say that there seems to be a reliable aspect of language proficiency closely linked to literacy, and other verbal-academic activities that are decontextualized, i.e. CALP (Cummins, 1980: 86). This dimension of CALP seems to be “largely independent of these language proficiencies which manifest themselves in everyday interpersonal communicative contexts” (Cummins, 1980: 86). Cummins (1980) clarifies that it has previously been hypothesized that the cognitive/academic aspects of L1 and L2 are interdependent, and that the development of L2 proficiency is partly an outcome of the L1 proficiency level at the onset of intensive L2 exposure. As Cummins (1980) quotes: “In other words, previous learning of literacy-related functions of language (in L1) will predict future learning of these functions (in L2)” (Cummins, 1980: 86). For this particular concept of L1 and L2 proficiency, Cummins (1980) adopts the ‘dual-iceberg diagram’ (see Figure 2). This diagram displays the presence of a common underlying, interdependent cognitive/academic language proficiency in both the L1 and L2, besides the surface features, e.g. the BICS, of both the L1 and L2. These surface features consist of phonology, syntax, and lexicon. As well as the surface features, the ‘dual-iceberg’ diagram allows for non-surface, non-interdependent proficiency features of both the L1 and the L2 possibly unrelated to CALP.

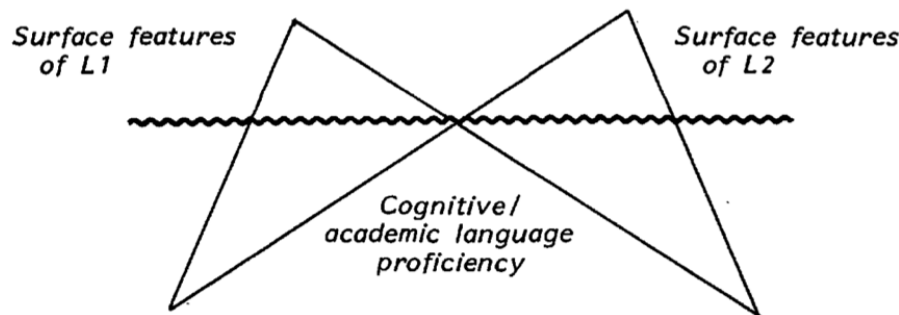


Figure 2: The 'Dual-Iceberg' Representation of L1 and L2 Language Proficiency (see Cummins, 1980: 87).

Although Cummins' theory regarding BICS and CALP refer to the role of language in academic achievement and the degree of active cognitive participation in the task or operation – particularly when it comes to bilingual speakers (Starfield, 1990: 84) – the model is still interesting for the current thesis. It could namely provide an explanation for the reason that both BICS, such as vocabulary, and CALP, such as reading ability, may manifest itself in the L2. If L1 and L2 CALP are related by a common underlying language proficiency, one might expect that L2 CALP will be significantly linked to L1 CALP measures, and there will be “a similar pattern of correlations with other variables such as verbal and nonverbal ability” (Cummins, 1980: 88). According to Krashen (1981), this underlying ability enables students, reacting to a new language, to demonstrate some kind of general understanding, and to make sense of the unfamiliar (Krashen, 1981: 159). Along these lines, students seem to transfer skills across languages, and as Krashen (1981) evinces, “this transfer is more likely to be observed among older students and among students with solid first language skills” (Krashen, 1981: 159). The argument that familiarity with either language will promote both L1 and L2 proficiency, provides sufficient motivation, and exposure both at school and in the wider context (Cummins, 2005: 5).

2.1.2 Krashen's Monitor Model

Precedently, the concept of transfer across languages according to Krashen (1981) was already mentioned. Krashen (1981) postulates that transfer takes place when the elements that were learned in the current task or skill are identical to those in the previously acquired task or skill (Krashen, 1981: 160). For instance, one might assume that language transfer with regard to the

skill of reading is found in the construct of CALP, as this construct covers literacy in general (Krashen, 1981: 163). The theories by Krashen (1981) are not solely concerned with an L1 influencing the output of an L2 with regard to BICS and CALP, but some notorious theories by Krashen are aimed at the proposal of one particular model: the Monitor model. Krashen (1978) argues that second language users often feel pressured to use correct and accurate language. However, an overconcern with correctness may be an issue as some language users are so concerned with form that they are not able to talk fluently at all (Krashen, 1978: 179). Therefore, a good understanding of language rules can be a real advantage for the language learner (Krashen, 1978: 179). Krashen (1980) proposes a model that accounts for some second language users demonstrating diverging performance in different situations. The elaboration of this model is essential in order to account for the finding that some students perform poorly on structure tests, while seeming to be able to interact, or communicate, in the target language quite well (Krashen, 1980: 213). As Blanche and Merino (1989) argue, the model proposes that while learning an L2, adult learners synchronously develop two potential independent structures regarding second language performance: one ‘acquired’, which is developed in ways similar to the learning of the L1, and the other ‘learned’, which is developed actively, or consciously, and predominantly in formal situations (Blanche & Merino, 1989: 326). This entails that the consciously acquired L2 system enables adult linguistic L2 production, with the naturally acquired system functioning only as a Monitor. The Monitor inspects and occasionally alters the output of the acquired system, when circumstances permit (Krashen, 1980: 213). The model assumes that performers may differ in the degree to which conscious monitoring is applied:

“At one extreme, there are performers who seem to monitor whenever possible (‘Monitor over-users’) and whose performance is therefore quite uneven. At the other extreme are performers who do not seem to use a Monitor at all (‘Monitor under-users’), even when conditions would allow it” (Blanche & Merino, 1989: 326).

The likelihood of foreign-language learners using the Monitor depends on the essence of the to be performed linguistic task and the emphasis that this task requires in comparison with other variables: “Tasks that cause students to focus on linguistic analysis (such as *fill in the blank with the correct morpheme*) would seem to invite monitoring, while tasks that impel the speaker to focus on communication (such as *answering a real question*) do not” (Krashen, 1982 in Blanche & Merino, 1989: 330). However, it is only possible to use the Monitor when

the following three requirements are met (1) time – to think about and use conscious rules; (2) ‘focus on form’ – to focus on form, or think about correctness; and (3) knowledge of the rule (Krashen, 1982: 16). Yet, in Krashen (1982) it is stated that: “[...] for most people, even university students, it takes a real discrete-point grammar-type test to meet all three conditions for Monitor use [...]” (Krashen, 1982: 18). Gregg (1984) argues that this implies that the Monitor cannot be used under normal circumstances. The conscious knowledge of rules (‘learning’) is only enabled by the use of the Monitor; therefore, the conscious knowledge of the rules is of little benefit when it comes to language *acquisition* (Gregg, 1984: 84). For language *learning*, however, the Monitor model’s theory might be of use. Gregg (1984) entails that language learning does not necessarily have to result in language acquisition for some language rules. As Krashen (1979) argues, some late-acquired rules, “[...] such as the third person singular ending on regular verbs in the present tense [...]” often seem to contain output errors in the utterances of the ESL performer (Krashen, 1979: 157). This could manifest itself although these performers have demonstrated to be excellent Monitor users (Krashen, 1979: 157). In sub-section 2.2.3, it will be clarified why it is important to have elaborated on the Monitor model, as the model may influence the outcomes of self-assessment.

2.1.3 Vocabulary Knowledge according to CEFR Language Standards

In the literature, when vocabulary knowledge is discussed, the terms *vocabulary size*, *vocabulary range*, and *vocabulary control* are often referred to. With regard to these contrasting notions, vocabulary size is often referred to as ‘lexical breadth’. According to Daller, Milton and Treffers-Daller (2007), vocabulary size is “the number of words a learner knows regardless of how well he or she knows them” (Daller, Milton, & Treffers-Daller, 2007: 7). According to Hulstijn (2007), vocabulary size is part of conscious knowledge of higher-order cognition (Hulstijn, 2007: 4). Regarding vocabulary range, Milton (2010) states that “much of the Vocabulary range criterion, with its characterizations of basic vocabulary and broad lexical repertoire appears to be a function of this size or breadth dimension” (Milton, 2010: 219). Vocabulary range, according to Milton (2010), is broadly a function of vocabulary size as well (Milton, 2010: 219). Vocabulary control is the ability to select the accurate word for the intended semantic meaning. Ho and Huong (2011: 15) argue that vocabulary knowledge plays an important role in the acquisition of English. The importance of vocabulary may manifest itself when learners read a text or communicate with another person, and come across words that are foreign to them, and that they do not understand (Laufer & Girsai, 2008). Ho and Huong (2011) quote the studies by Huckin and Bloch (1993) and Nation (1994) by stating that learners

rely on vocabulary as their primary resource, and a rich vocabulary promotes listening, speaking, reading, and writing skills (Huckin & Bloch, 1993; Nation, 1994 in Ho & Huong, 2011: 15-16). English vocabulary, in particular, has been reported to be one of the most difficult, hence English as a foreign language (EFL) learners face a common problem of vocabulary insufficiency (Ho & Huong, 2011: 23). Furthermore, the authors of the latter article demonstrate that spelling is one of the abilities that the learners perform the worst on, which has a negative effect on writing skills (Ho & Huong, 2011: 23). Janulevičienė and Kavaliauskienė (2011), likewise, argue that vocabulary knowledge is an important predictor of language acquisition skills; therefore, it could be argued that language learners make every effort to achieve perfection in this linguistic domain (Janulevičienė & Kavaliauskienė, 2011: 11). However, perfecting vocabulary use may be a challenge, as it involves the handling of multiple meanings. Janulevičienė and Kavaliauskienė (2011) provide the following examples of (semantic) meanings, that language learners have to deal with: “[...] propositional meaning, register, metaphorical meaning, connotational meaning and the representation of meaning such as definition, relationships – synonymy, autonomy, hyponymy, meronymy, collocation, translation, etc.” (Janulevičienė & Kavaliauskienė, 2011: 11). It might, therefore, be quite difficult to accurately assess the vocabulary of the second language learner.

With regard to vocabulary knowledge, the Common European Framework of Reference (CEFR) provides descriptors that enable to classify learners in six different language proficiency levels: A1, A2, B1, B2, C1, and C2. Here, the levels are categorized from the lowest level to the highest level. Faez, Majhanovich, Taylor, Smith and Crowley (2011) postulate that the CEFR’s descriptors are written in the form of ‘Can Do’ statements, for each category, explaining what learners should be able to do at each L2-proficiency level. One important notion Faez et al. (2011) emphasize, is that the CEFR is rather descriptive than prescriptive, hence the CEFR does not recommend any specific teaching or assessment methods (Faez et al., 2011: 5). According to Milton (2010), the CEFR hierarchy implies that development through hierarchy is closely related to the knowledge of vocabulary, and to learning more foreign language words. High level performers tend to have extensive knowledge of vocabulary, while performers at the primary stage do not (Milton, 2010: 218). Furthermore, Milton (2010: 218) evinces that knowledge of the most common and frequent words in the foreign language seems to be essential for good output performance. Likewise, according to Milton (2010), the importance of the CEFR lies in the ability of its users to apply the criteria on the descriptors correctly and reliably. However, this can be difficult to implement in practice in the absence of

more detailed criteria. The CEFR indirectly identifies this problem by suggesting that it may be useful to add specifics of the vocabulary to the descriptors (Milton, 2010: 214). The CEFR-level descriptors also provide reference to the vocabulary that may be required of learners practicing certain *skills* (Milton, 2010: 213). These references could for instance be the identification and comprehension of familiar words in the A1 listening and reading descriptors, and in B1 reading descriptors it could entail reference to the comprehension of “high frequency or everyday job-related vocabulary” (Milton, 2010: 213). Hulstijn (2007) argues that the concept of language proficiency provided in the CEFR rests on two foundations (quantity and quality), which are strongly interwoven. Quantity elements determine what the learner can do and quality elements show how well the learner is able to do this (Hulstijn, 2007: 2). Hulstijn (2007) quotes the study by De Jong (2004), in which it is argued that quantity refers to the number of domains, functions, notions, circumstances, places, topics, and positions that a language user may tackle. Quality refers to the degree of accuracy and effectiveness in both understanding and expression of meaning, and the degree to which language use is efficient, leading to communication with the least effort possible (Hulstijn, 2007: 2). If a learner is placed at an overall production level of B2, it is not certain that that learner must also have achieved the B2 level scales of all the linguistic abilities, or whether a learner can be placed at different levels of different scales (Hulstijn, 2007: 2). In the research of the latter author, it is stated that there are three types of L2 users: users with low linguistic quantity abilities but high linguistic quality abilities; users with high linguistic quantity abilities but low linguistic quality abilities; and users with matched quantity and quality linguistic abilities, as proposed by mixed type CEFR scales (Hulstijn, 2007: 3). Apart from the quality and quantity scales of the CEFR, scales of a variety of linguistic competences are covered by the framework as well. These competences may be “Vocabulary Range, Vocabulary Control, Grammatical Accuracy, and Phonological Control” (Hulstijn, 2007: 2).

Jones and Saville (2009) argue that The Council of Europe’s activities aim to promote linguistic diversity and language learning (Jones & Saville, 2009: 52). Although the CEFR is beneficial for European citizens, its empirical foundation consists of observations of language teachers and other experts on descriptor scaling (Hulstijn, 2007: 7). This implies that teachers often assign one student as a reference point, determining whether or not that student was able to do what was specified in the descriptors. Nonetheless, the CEFR is not based on empirical evidence derived from data from L2 learners (Hulstijn, 2007: 7). North and Schneider (1998) even argue that CEFR scales can be seen as unidimensional when considering the

psychometrics behind it (North & Schneider, 1998: 238-239 in Hulstijn, 2007: 7). Hulstijn (2007) states that the CEFR poses some challenges, as it is not empirically established that all L2 students at any other functional level than A1, say B2, reach that level by passing the level below; i.e. B1 in this case. Furthermore, there has not been provided any empirical evidence to suggest that all L2 learners at a given level are capable of performing all the tasks associated with lower levels. The latter, undoubtedly, does not apply to the lowest level A1, as in this case there are no tasks associated with a lower level. The fact that L2 learners should gradually move into the next scale, and are able to perform tasks associated with the lower levels should, however, be accounted for in order for the CEFR scales to be truly implicit and unidimensional (Hulstijn, 2007: 8). One last challenge with the credibility of the CEFR scales, according to Hulstijn (2007), is that there is no evidence that a learner at a particular level of an overall scale, such as B2 Overall Oral Production, automatically has the same consistency in terms of the other language scales; “[...] e.g. B2 Vocabulary Selection, B2 Grammatical Precision, and B2 Phonological Control [...]” (Hulstijn, 2007: 8). According to Milton (2010), the potential benefit of a method of assessment that can provide the CEFR characteristics with certain accurate measurements is quite evident. This implies that when a learner knows and can apply several thousands of words, idiomatic and colloquial phrases included, and his or her foreign language vocabulary comprehension is equivalent to a native speaker, there exists strong evidence that he or she has C2 level proficiency, “[...] at least in terms of vocabulary range” (Milton, 2010: 214). If a learner solely knows and can apply a few hundred foreign language words, it is likely that the learner is placed at A1 level proficiency in terms of vocabulary range (Milton, 2010: 214).

2.2 The Correlation between Standardized Assessment and Self-Assessment

This section consists of three sub-sections. The first section looks into the assessment of vocabulary by means of standardized assessment, which is considered a valid instrument to determine actual language proficiency. The second section provides some insights on the concepts of self-assessment, and in the third section, the reader is provided with an elaboration on particular insights that clarify the affiliation of the reported estimated level of language proficiency and standardized assessment.

2.2.1 Assessing Vocabulary Knowledge with Standardized Assessment

Anderson (1998) notes that assessment fulfills an essential role. Amongst others, it impacts what is taught and learned. As Hughes (1989) states, knowledge about one's language proficiency is often "very useful and sometimes necessary" (Hughes, 1989: 4). In order to chart language proficiency in a standardized manner, multiple assessments could be employed. Hughes (1989) argues that vocabulary testing is designed to obtain insights in general indications about the adequacy of the learners' vocabulary. Therefore, a published vocabulary test could be employed (Hughes, 1989: 179). Pearson, Hiebert and Kamil (2007) provide a clear overview of the onset of vocabulary assessment, and what vocabulary assessments actually measure. They state that the assessment of the learners' comprehension of word meanings, i.e. vocabulary assessment, is as old as the assessment of reading ability; so much as it could be argued that vocabulary assessment occurred in early tests of intelligence that preceded formal measures of reading comprehension (Pearson, Hiebert & Kamil, 2007: 285). It is likely that the onset of vocabulary assessment lies in measurements such as asking students to define or explain selected words as they were likely to be enclosed in texts that they encounter in their school work (Pearson, Hiebert & Kamil, 2007: 285). However, it is quite evident that this was time-consuming, and there should be a valid and reliable assessment method that ensures that all learners could be assessed in the same manner. Pearson, Hiebert and Kamil (2007) quote the study by Resnick and Resnick (1977), in which it is stated that the need for more effective, easier-to-administer, and easily scorable assessment was triggered through the drive toward mass testing, as there was a need to assess recruits for World War I (Resnick & Resnick, 1977 in Pearson, Hiebert & Kamil, 2007: 285). Therefore, standardized assessment with multiple-choice formats was employed. Read (2000) argues that, until the 1970s, the multiple choice format was the item type that occurred most in vocabulary assessment (Read, 2000 in Pearson, Hiebert & Kamil, 2007: 285). Thereafter, more contextualized vocabulary tests were developed as advances in language learning arose from the emerging fields of psycholinguistics, and cognitive science (Pearson, Hiebert & Kamil, 2007: 285).

In the handbook composed by Read (2000), there are three continua that could be identified while designing and evaluating vocabulary assessments. These continua could be identified in existing tests, and they are adjusted by Pearson, Hiebert and Kamil (2007). This results in the elaboration on the following continua: (1) discrete-embedded, (2) selective-comprehensive, and (3) contextualized-decontextualized (Read, 2000 in Pearson, Hiebert & Kamil, 2007: 287). The first continuum, discrete-embedded, distinguishes between the concept of considering

vocabulary as a separate construct with its own collection of test items, and its own performance report (i.e. the discrete end of the continuum), and the concept of considering vocabulary as an embedded construct that leads to – but is not considered separate from – a wider comprehension of the text (Pearson, Hiebert & Kamil, 2007: 287). The second continuum, the selective-comprehensive, distinguishes the relation between the sample of items in an assessment, and the hypothetical population of vocabulary items identified by the sample. As demonstrated by Pearson, Hiebert and Kamil (2007), this implies the following: “In general, the smaller the set of words about which we wish to make a claim, the more selective the assessment” (Pearson, Hiebert & Kamil, 2007: 288). The third – and last in the adjusted version by Pearson, Hiebert and Kamil (2007) – continuum is the contextualized-decontextualized continuum. This continuum distinguishes between vocabulary tests that differ in the extent to which textual context is necessary to determine the meaning of a word. Pearson, Hiebert and Kamil (2007) evince that any word in decontextualized format can be readily and easily evaluated. Yet, merely interpreting a word in a contextualized format does not automatically indicate the meaning is required to evaluate its significance: “In order to meet the standard of assessing students’ ability to use context to identify word meaning, context must actually be used in completing the item” (Pearson, Hiebert & Kamil, 2007: 289).

In most published tests, where vocabulary is tested by means of multiple choice items, the ability that is tested is *recognition ability*. The reason why multiple choice items are often employed, is that the “[...] distractors are usually readily available”, and “[...] there seems unlikely to be any serious harmful backwash effect, since guessing the meaning of vocabulary items is something that we would probably wish to encourage” (Hughes, 1989: 180). Two examples of item operations that Hughes (1989) provides, are the recognition of synonyms, and the recognition of definitions. The words that are enclosed in the sample of an assessment can be arranged according to their frequency and convenience. Hughes (1989) quotes: “From each of these groups, items can be taken at random, with more being selected from the groups containing the more frequent and useful words” (Hughes, 1989: 180). Likewise, the vocabulary that is employed in a standardized test could, according to Read (2007) be obtained by means of word frequency lists. This means examining the extensive literature and generated computer data on the vocabulary size of English native speakers (Read, 2007: 107). Nonetheless, Read (2007) argues that there does not exist a definitive word frequency list, either for English in general or for particular English uses (Read, 2007: 109). Assessments with regard to vocabulary size for second language learners are focused on a narrower range of words than those for native

speakers, as low-frequency words are far less likely to be understood, particularly by foreign language learners (Read, 2007: 108). When an appropriate word frequency list has been selected, choosing a set of target words for the test items is the next step in creating a vocabulary size assessment. In order to make accurate vocabulary size estimates, test designers appear to prefer using a simple test format since a fairly large sample is needed (Read, 2007: 110). As Read (2007) evinces, these test formats could for instance be multiple choice matching words formats with synonyms or brief word meanings. The latter item formats provide direct evidence that the test taker knows a particular word meaning (Read, 2007: 110). Another format Read (2007) clarifies, is the Yes/No format, with which test takers are given a set of words, and they are required to indicate whether they know a particular word or not (Read, 2007: 110). Regarding the assessment of quality (or depth) of vocabulary knowledge, Read (2007) argues the following:

“[...] there is in fact much more to know about words if they are to become functional units in the learner’s L2 lexicon: how the word is pronounced and spelled, what its morphological forms are, how it functions syntactically, how frequent it is, how it is used appropriately from a sociolinguistic perspective, and so on” (Read, 2007: 113).

This entails that a vocabulary size test, which usually tests if the learner is able to associate the written form of a word with an elementary assertion of its semantic meaning, might not be sufficient to chart the entire volume of one’s vocabulary knowledge (Read, 2007: 113).

2.2.2 Delineating the Concept of Self-Assessment

As learners assess their own abilities, and aggregate their language learning experiences, self-assessment instruments will provide ample proof of individual, and collective achievement (Brantmeier, Vanderplank & Strube, 2012: 145). At that time of writing, Blanche and Merino (1989) stated that the topic of self-assessment had just begun to grow as a distinct field of interest in the field of language testing (Blanche & Merino, 1989: 315). Self-assessment is the capacity to appraise one’s own performance (Baleghizadeh & Masoun, 2014: 27). Self-assessment can also be described as “information about the learners provided by the learners themselves, about their abilities, the progress they think they are making, and what they think they can do or cannot do yet with what they have learned in a course” (Blanche & Merino, 1989 in Baleghizadeh & Masoun, 2014: 27). Self-assessment has been identified as an important method as it also gives learners the opportunity to participate in the learning process by

evaluating their own strengths and weaknesses (Baleghizadeh & Masoun, 2014: 27). As Dandenault (1997) states, the term ‘self-assessment’ is often interchangeably referred to as ‘self-rating’, ‘self-evaluation’, ‘self-control’, or ‘self-appraisal’. However, according to Blanche and Merino (1989), the term ‘self-assessment’ is preferred in the literature. In Dandenault (1997), it is stated that the latter term is “a less loaded term in that it does not carry such a final and evaluative connotation” (Oscarson, 1980 in Dandenault, 1997: 3). Self-assessment in the current study refers to subjects assessing their own performance according to their intuitions, and the focus lies on the abilities the participants think they possess. As Nurov (2008) demonstrates, self-assessment practices can be pre- or post-facto. Pre-facto self-assessment is conducted prior to an external evaluation, such as a standardized test or a teacher’s review assessment (Nurov, 2000: 18). Self-assessment after an external test is performed, is called ‘post-facto self-assessment’. Nurov (2000) quotes the study by Blanche (1990) in which a method of self-assessment post-facto procedure is applied. In this study, the participants were asked to evaluate their performance on a standardized test after they took it (Nurov, 2000: 19). Not only can self-assessment practices be divided in pre-facto and post-facto; they can also be classified into norm-referenced and criterion-referenced self-assessments. In a norm-referenced self-assessment, learners assess their skills compared to others, most often using concise general skill benchmarks (Nurov, 2000: 19). Criterion self-assessments allow learners to evaluate themselves against clear requirements or criteria (e.g. course objectives) (Nurov, 2000: 19).

LeBlanc and Painchaud (1985) initially argued that students do not seem to have the tools to manage the task of precisely self-assessing themselves regarding their level of proficiency (LeBlanc & Painchaud, 1985: 675). Nonetheless, MacIntyre, Noels and Clement (1997) propose that language learners should be able to accurately rate their own abilities (MacIntyre, Noels & Clement, 1997: 267). As Nurov (2000) states, the self-assessment approach is based on the assumption that learners can most accurately evaluate themselves, because they have access to a broad database of their own achievements and skill deficiencies (Nurov, 2000: 8). According to Sedikides and Strube (1997), people, in general, are encouraged to achieve a consensually correct self-evaluation according to the self-assessment perspective. To achieve this goal, people are primarily interested in the diagnosticity of self-relevant knowledge; i.e. the degree to which this knowledge can reduce uncertainty about an aspect of the self (Sedikides & Strube, 1997: 213). Sedikides (1993) demonstrates that diagnostic tests or tasks contribute to a correct self-image, and that these tasks possess high informative value, with which it can

clearly differentiate between people with high-level ability and people with low-level ability (Sedikides, 1993: 317). Sedikides and Strube (1997) evince that people seek diagnostic information regardless of its positive or negative self-impacts, and whether the information confirms or contradicts current self-conceptions. Self-assessment, in short, serves the purpose of increasing the certainty that holds self-knowledge (Sedikides & Strube, 1997: 213). Likewise, Sedikides (1993) argues that people are driven to minimize uncertainty about their abilities or personality traits according to the self-assessment perspective. In self-evaluative environments, ambiguity is minimized by creating an objective and reliable representation of the self (Sedikides, 1993: 317). As previously mentioned, according to Moritz (1996) self-assessment is “[...] influenced by individual experiences and language learning backgrounds, as well as individually-determined strategies for approaching the self-assessment task, individually-defined points of comparison [...], and individual levels of self-confidence, both with regard to foreign language abilities, and to answers on the questionnaire” (Moritz, 1996: 17). According to Blanche (1988), foreign language learners can be at a disadvantage with regard to self-assessment accuracy being a condition of learner autonomy, as they are often not able to equate themselves with native speakers. In addition, the reliability of their judgements can be impeded by the fact that language learning is a dynamic process which is closely connected to subjective factors, such as personality characteristics (Blanche, 1988: 75).

As previously mentioned, self-assessment is influenced by self-confidence. What is more, individual characteristics such as an integrative motive, or self-confidence with English have been aspects hypothesized to affect actual language competence (Clément, Gardner & Smythe, 1980: 294). In this sense, self-confidence influences both self-assessment and language competence. In the context of their study, Clément, Gardner and Smythe (1980) report that self-confidence in English tends to derive from the actual use of language by individuals outside of school and at home. Individuals who frequently have interactions with Anglophones will develop self-confidence in themselves with their English skills, will be inspired to learn English, and will be fairly proficient. Therefore, personal interaction seems to be an important factor in building self-confidence in English (Clément, Gardner & Smythe, 1980: 299). Self-confident learners are most likely to possess a self-enhancing bias (MacIntyre, Noels & Clement, 1997: 269). As Taylor and Brown (1988) state, self-enhancement aids in the development of new skills, as it provides the force to spend the extra effort required to tackle a linguistic challenge. A positive bias may in fact support the language learning process by increasing the ability of the student to interact in the L2 and facilitate language learning

(MacIntyre, Noels & Clement, 1997: 279). One important approach is that not believing in the ability to learn or perform in an L2 creates negative expectations which in turn lead to decreased effort and accomplishment (MacIntyre, Noels & Clement, 1997: 280). Moreover, several social psychological motivation models indicate that expectations of performance interfere between real competency and subsequent achievement. MacIntyre, Noels and Clement (1997) cite the studies by Bandura (1986, 1988) in which it is clarified that: “If expectancies [of competence] are high, then one will expend greater effort, with greater likelihood of success. If, on the other hand, expectancies are low, one expends less effort, with concomitantly less success” (MacIntyre, Noels & Clement, 1997: 267). Moritz (1996) evinces that there has previously been provided some evidence that self-esteem, and other psychological factors may play a role in how learners assess their abilities (Moritz, 1996: 3-4). Keeping in mind that the experiment in the current thesis is performed with Dutch participants, the fact that self-esteem might influence the outcomes of self-assessment is something to consider. That is, Dutch society is perceived as highly individualistic (Oppenheimer, 2004: 337), and individualism is related to high self-esteem (Verkuyten, 2009: 424).

2.2.3 Self-Assessment and its Affiliation to Standardized Assessment

In the previous section (see section 2.1.2), some background information regarding Krashen’s Monitor theory was elaborated on. It was already stated that the Monitor may influence the accuracy of self-assessment. According to Blanche and Merino (1989), this could be determined by the fact that while the Monitor only has the function to inspect and (occasionally) alter, users of the Monitor often self-correct using *acquisition* in both the L1 and the L2 (Krashen, 1982 in Blanche & Merino, 1989: 326). As argued by Blanche (1988), this entails that researchers should first attempt to decide if their participants are more likely to use the Monitor or the system they have acquired for self-assessment. In turn, legitimately comparing certain assessment outcomes would be more straightforward (Blanche, 1988: 83).

In the investigation by Alderson (2005), the self-assessment component of DIALANG was employed in the experiment. According to Brantmeier and Vanderplank (2008), in low-stake testing settings such as the DIALANG test, learners should be able to compare their self-assessment ratings with their performance in any specific skill so that the differences exposed will provide useful insights into their language learning and beliefs (Brantmeier & Vanderplank, 2008: 459). Brantmeier and Vanderplank (2008) report that the results of the investigation by Alderson (2005) revealed that there was a significant relationship between self-

assessed reading level, and levels of items linked to the Common European Framework (CEFR) regarding reading ability. There were also notable differences in demographic variables such as mother tongue, age, gender, length of time learning English and frequency of use (Brantmeier & Vanderplank, 2008: 459). Although Blanche (1988) argues that self-assessment activities seem to improve the enthusiasm of the learners, Baleghizadeh and Masoun (2014) postulate that there are certain drawbacks when it comes to applying self-assessment. Leach (2012) investigated self-assessment accuracy, and reported that high attainers tended to underestimate, and low attainers tended to overestimate their performance (Leach, 2012 in Baleghizadeh & Masoun, 2014: 28). Similarly, according to Blanche (1988), it has previously been shown in various studies that more capable students appear to underestimate their language abilities. Alternatively, poor students tend to overestimate to a greater extent (Blanche, 1988: 82). Kruger and Dunning (1999) as well state that whereas low-performing learners frequently overestimate their skills, high-performing learners tend to underestimate their performance. Furthermore, according to Falchikov and Boud (1989), more experienced students are in favor of common sense predictions about their performance, as these students tend towards greater accuracy in their ratings than less experienced students. Even the more, experienced students tended to undervalue their results as well. Alderson (2005), likewise, evinces that higher attaining learners were more able to self-assess in their pilot English experiments than lower attaining learners (Alderson, 2005 in Brantmeier & Vanderplank, 2008: 459). According to Brantmeier and Vanderplank (2008), one potential explanation for the discrepancies in these self-assessments is that students can differentiate between what they consider their 'real life' performance to be in a foreign language and what they accomplish in tests (Brantmeier & Vanderplank, 2008: 471-472). However, as Falchikov and Boud (1989) report, there is no overall consistent tendency to over- or underestimate performance (Falchikov & Boud, 1989: 396). Nurov (2000) cites the study by Hughes (1989) by stating that one drawback with regard to self-assessment is that it is the least objective of all human behavior measures (Hughes, 1989 in Nurov, 2000: 25). As Hughes (1989) demonstrates, an objective measure is not influenced by personal judgement and decision. Test objectivity relates to its method of scoring (Hughes, 1989), with the scoring of an objective test not being determined by personal judgement (Bachman, 1990; Hughes, 1989 in Nurov, 2000: 25). Nurov (2000) quotes the research by Brown (1996), in which it is revealed that standardized assessments consisting of, for instance, multiple-choice items are more objective, as the correct answers to these items and the scoring criteria are predetermined and are not dependent on subjective human decisions and feelings (Brown, 1996 in Nurov, 2000: 25). Self-assessment is subjective as it is solely based on

personal opinion and judgment (Nurov, 2000: 25). However, objectivity of scoring should not be confused with validity and reliability, as objectivity does not undoubtedly entail validity and reliability (Nurov, 2000: 26).

While standardized assessment is preferred as it is a reliability requirement, which is a required prerequisite for validity, contextualized and non-standardized (alternative) assessment, such as self-assessment can say more about the skill of the learner than decontextualized assessments (Moss, 1994 in Nurov, 2000: 24). Self-assessment can be more interpretative of the language behavior of the learner than the standardized test, as it does not only represent the degree particular behavior was produced but also how the behavior evolved (Rea-Dickins & Germaine, 1996; Tudor, 1996 in Nurov, 2000: 24-25). As reported by Bailey (1998), most self-assessment research has looked into validity criteria: “Criterion validity is the extent to which an assessment instrument agrees with other measuring instruments whose validity is thought to have been established (Bailey, 1998; Hughes, 1989)” (Nurov, 2000: 25). The criterion validity of self-assessment can be calculated by measuring the degree of its consistency with external criteria such as standardized tests and evaluation of teachers, given that they evaluate the same construct; i.e. the same characteristic or skill (Nurov, 2000: 25). According to Moritz (1996), research on self-assessment in foreign language pedagogy has concentrated predominantly on concurrent or criterion validity, and generally yielded unimpressive results. Most studies compare the outcomes of self-assessments “[...] with either: a) the results of a previously-established, ‘objective’ test, b) a teacher’s ratings of a student on the same scale, or c) a final course grade” (Moritz, 1996: 3). The majority of these studies report their statistics by means of Pearson Product-Moment correlations. Nurov (2000) quotes a study by Buck (1992) that shows that a fairly large number of EFL-learners in a variety of Japanese universities and colleges in Osaka have been able to self-assess their listening and reading skills to a reasonably positive degree of interaction with external assessment procedures ($r > .50$ in all correlations) (Nurov, 2000: 34-35). Nonetheless, Nurov (2000) states that in the study by Blanche and Merino (1989), it was reported that the majority of self-assessment research provides strong and credible proof of self-assessment reliability and validity (Nurov, 2000: 36). Correlation values of self-assessment with objective testing tools such as standardized assessments and language ability estimates on the part of teachers ranging from .50 to .60 are normal with higher correlations not being implausible (Nurov, 2000: 36-37). However, almost every researcher who reports about an empirical self-assessment inquiry seems to have a different understanding

of what the correlation rates actually mean. Regarding the correlations between self-assessment and other assessment ratings, Moritz (1996) quotes the following:

“According to LeBlanc & Painchaud’s earlier articles (1980, 1981, 1982, and others), a correlation of .49 is ‘good evidence’ that students can self-assess, while Janssen-van Dielen (1989) calls correlations ranging from .29 to .69 ‘too low’. Likewise, Wesche, Morrison, Ready, and Pawley (1990) deem a correlation of .58 ‘quite low’, while Krausert (1991) argues that her correlations of .36 to .54 are ‘high’” (Moritz, 1996: 3-4).

It can be concluded that the perceptions of a significant correlation of self-assessment and other types of standardized assessment are severely divergent. Nonetheless, one matter could be claimed with certainty: “[...] self-assessment instruments yield higher correlations with measures of proficiency if the self-assessment items are specific and focused (Pierce et al., 1993)” (Brantmeier, Vanderplank & Strube, 2012: 152). With regard to the self-assessment of vocabulary knowledge, Janulevičienė and Kavaliauskienė (2011) state that vocabulary, and its use play a significant role in determining the individual linguistic ability of the learners. However, as Janulevičienė and Kavaliauskienė (2011) argue, the capacity of the learners to determine their own language competence and usage is not always unbiased. Quite frequently, learners misinterpret their capacity to express knowledge and ideas at “a different and higher than everyday-language level” (Janulevičienė & Kavaliauskienė, 2011: 14). In the latter article, the participants – 150 English for Legal Purposes (ELP) students at an intermediate level – quite frequently overestimated their vocabulary knowledge: “The key cause of linguistic deficit might be learners’ inability to internalize knowledge of ESP vocabulary, i.e. to transfer knowledge to its usage” (Janulevičienė & Kavaliauskienė, 2011: 14). Furthermore, the different types of word meaning, regarding vocabulary (elaborated in section 2.1.3) might result in poor correlations between self-assessment and actual performance (Janulevičienė & Kavaliauskienė, 2011: 11). One implication of the latter notion is that it might apply to some participants in the experiment of the current thesis as well. In other words, transferring knowledge to usage may cause complications in accurately self-estimating vocabulary knowledge.

3. Research Questions and Hypotheses Development

The current chapter starts out with the formation of the research questions. The preceding theoretical frame enables to subsequently form a set of hypotheses with regard to these research questions.

3.1 Research Questions

This study aims to report on one main research question. The report on this research question (RQ) is aided by the existence of a set of three sub-questions. Hence, with these sub-questions, it is aimed to answer the main RQ. The RQs and sub-questions may be formulated as follows:

RQ: Is there a significant correlation between CEFR levels of vocabulary knowledge, according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test, and a post-test self-assessment component?

Sub-questions:

1. Are there discrepancies in the correlations between the post-test self-assessment component as a construct, and the CEFR levels of vocabulary knowledge according to either the DIALANG Vocabulary Test or the Cambridge General English Online Test?
2. Are there discrepancies in the correlations between the two parts – the CEFR descriptors of vocabulary knowledge, and the Likert scale ratings concerning listening comprehension, oral proficiency, reading ability, and writing ability – of the post-test self-assessment component and the language testing component?
3. Are there demographic discrepancies, i.e. gender, educational level and use of English, regarding the correlation of the self-assessment component and the language testing component?

3.2 Hypotheses Development

Primarily, it is to be declared that the following alternative hypotheses are accepted as the valid assumptions regarding the research question and the sub-questions.

Cummins' 'iceberg' model indicates that there seem to be basic interpersonal communicative skills (BICS), that are directly perceivable, as they are 'above the surface' language features.

As previously mentioned, it could be argued that both BICS, such as vocabulary, and CALP, such as reading ability, may manifest itself in the L2. It is therefore expected that the participants in the current experiment are able to demonstrate their vocabulary knowledge. Krashen's (1980) Monitor model accounts for certain second language users showing varying output in different circumstances. To account for the finding that some students perform poorly on structure tests while appearing to be able to interact or communicate very well in the target language, the elaboration of this model has been shown to be important (Krashen, 1980: 213). The Monitor model may influence the accuracy of self-assessment. According to Blanche and Merino (1989), this could be determined by the fact that while the Monitor has only the function to inspect and (occasionally) alter the output, users of the Monitor often self-correct using both L1 and the L2 *acquisition* (Krashen, 1982 in Blanche & Merino, 1989: 326). Furthermore, in previous research by Alderson (2005), it was evinced that there exists a significant relationship between self-assessed reading level and levels of items linked to the Common European Framework (CEFR) regarding reading ability. It is therefore likely that this may also be the case for vocabulary knowledge. Moreover, the majority of self-assessment research provides strong and credible proof of self-assessment reliability and validity (Nurov, 2000: 36). Correlation values of self-assessment with objective testing tools such as standardized assessments and language ability estimates on the part of teachers ranging from .50 to .60 and in some cases even higher are apparent (Nurov, 2000: 36-37). Therefore, the following hypotheses with regard to the main research question could be formulated:

H_{0RQ}: There is no significant correlation between CEFR levels of vocabulary knowledge, according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test, and a post-test self-assessment component.

H_{ARQ}: There is a significant correlation between CEFR levels of vocabulary knowledge, according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test, and a post-test self-assessment component.

The formation of the hypotheses with regard to the three sub-questions will be specified in the following sections. The DIALANG Vocabulary Test adopts more open-ended questions (i.e. less multiple choice items) than the Cambridge General English Online Test. The latter solely contains multiple choice items. As previously mentioned, Hughes (1989) postulates that multiple choice items are a safe choice in language testing, as "there seems unlikely to be any

serious harmful backwash effect, since guessing the meaning of vocabulary items is something that we would probably wish to encourage” (Hughes, 1989: 180). Therefore, it is expected that the test takers score better on the Cambridge General English Online Test in general. This may result in the correlation between the test performance and the estimated proficiency level, obtained in the self-assessment part of the experiment, being different for each test. Hence, the following hypotheses regarding the first sub-question are formed:

H₀₁: There are no discrepancies in the correlations between the post-test self-assessment component as a construct, and the CEFR levels of vocabulary knowledge according to either the DIALANG Vocabulary Test or the Cambridge General English Online Test.

H_{A1}: There are discrepancies in the correlations between the post-test self-assessment component as a construct, and the CEFR levels of vocabulary knowledge according to either the DIALANG Vocabulary Test or the Cambridge General English Online Test.

As previously mentioned, the criterion validity of self-assessment can be calculated by measuring the degree of its consistency with a standardized test, given that they evaluate the same construct; i.e. the same characteristic or skill (Nurov, 2000: 25). The perceptions of what should be considered a significant correlation of self-assessment and other types of standardized assessment are divergent. Nonetheless, one matter could be claimed with certainty: “self-assessment instruments yield higher correlations with measures of proficiency if the self-assessment items are specific and focused (Pierce et al., 1993)” (Brantmeier, Vanderplank & Strube, 2012: 152). The self-assessment item types concerning estimates regarding listening comprehension, speaking, reading and writing skills are perhaps not evidently specific and focused on vocabulary knowledge when comparing them with the CEFR vocabulary knowledge descriptors. Therefore, it is expected that both these self-assessment features yield different results in the correlation between the self-assessment rates and the test outcomes. The following hypotheses could be formed:

H₀₂: There are no discrepancies in the correlations between the two parts – the CEFR descriptors of vocabulary knowledge, and the Likert scale ratings concerning listening comprehension, oral proficiency, reading ability, and writing ability – of the post-test self-assessment component and the language testing component.

H_{A2}: There are discrepancies in the correlations between the two parts – the CEFR descriptors of vocabulary knowledge, and the Likert scale ratings concerning listening comprehension, oral proficiency, reading ability, and writing ability – of the post-test self-assessment component and the language testing component.

Clément, Gardner and Smythe (1980) state that frequently interacting with Anglophones will develop self-confidence in the learners' English skills, and the learners will be inspired to learn English, and will be fairly proficient. Therefore, personal interaction in the language seems to be an important factor in building self-confidence in English (Clément, Gardner & Smythe, 1980: 299). More frequently conversing in English could therefore be a factor that results in higher test performance in the present experiment. Furthermore, in the study by Alderson (2005), there were notable differences in demographic variables such as mother tongue, age, gender, length of time learning English, and frequency of use in the relationship between self-assessment and reading ability according to the CEFR (Brantmeier & Vanderplank, 2008: 459). Therefore, it is to be expected that the current experiment yields demographic discrepancies in the correlation of self-assessment ratings and the test scores as well. This results in the formation of the following hypotheses:

H₀₃: There are no demographic discrepancies, i.e. gender, educational level and use of English, regarding the correlation of the self-assessment component and the language testing component.

H_{A3}: There are demographic discrepancies, i.e. gender, educational level and use of English, regarding the correlation of the self-assessment component and the language testing component.

4. Methodology

This chapter will provide the reader with an outline of the research methodology. This includes an elaboration of the participants that participate in this study, the instrumentation of the material that was developed for the experiment, a description of the procedure that was employed during the experiment, the research design and the qualitative and statistical analyses that will be performed in the current research.

4.1 Participants

The participants in this study consist of 73 persons with the Dutch nationality and Dutch as a native language. This could be ascertained in advance, as the main language of the questionnaire is Dutch. Age ranges from 20 to 67 with a mean age of 34 ($M = 34$, $SD = 13$). The sampling method with which the participants are recruited is a non-probability sampling method¹, where (1) convenience sampling, (2) voluntary response sampling, and (3) snowball sampling are simultaneously applied. That is, some participants are recruited via convenience sampling, as they have been directly requested by the researcher to participate in the experiment. Moreover, as the questionnaire was posted online, some participants chose to voluntarily participate in the experiment. Lastly, some participants requested other participants to participate in the experiment. Therefore, a snowball sampling method does apply as well. However, the participants are not necessarily selected based on non-random criteria, as no specific criteria are concerned in the recruitment of the participants. Furthermore, every individual in the entire population does have a chance of being included in the current sample, considering the questionnaire is publicly spread via social media.

The following section will provide a brief outline of some demographical features of the participants. Of the 73 participants that participated in the study, there are 57 female participants (78,1%) and 16 male participants (21,9%). One of the participants (1,4%) is bilingual, with both Dutch and English as a native language, and the remainder of the participants (98,6%) are speakers of Dutch as a native language. With regard to educational background², the majority of the participants has graduated HBO ($n = 23$), followed by the participants that graduated MBO ($n = 12$) and participants that have a WO-diploma ($n = 11$). Part of the participants are still employed in HBO ($n = 10$) and university ($n = 9$), or MBO ($n = 1$). For few participants, the highest level of education is secondary school ($n = 5$), and a small amount of the participants would rather not declare their educational background ($n = 2$). With regard to use of English, two questions in the questionnaire are developed to indicate this matter: (1) use of English in daily life, and (2) use of English in the workplace or in the educational setting. The majority of participants ($n = 28$) indicated that they sometimes use the English language in their daily life, followed by the participants that do not use English in their daily life ($n = 23$). The remainder of the participants indicated to use English in their daily life ($n = 22$). Regarding the use of

¹ For further reference, see <<https://www.scribbr.com/methodology/sampling-methods/>>.

² See Appendix 9.3 for a further elaboration on the educational system in The Netherlands.

English in the workplace or educational setting, the majority of participants ($n = 26$) indicated that they sometimes use the English language in their workplace or educational setting, followed by the participants that do not use English in their workplace or educational setting ($n = 24$). The remainder of the participants indicated to use English in their workplace or educational setting ($n = 23$).

4.2 Instrumentation

The experiment of this study consists of two main components: (1) the language testing component, and (2) the self-assessment component. The first component employs both the DIALANG Vocabulary Test and the Cambridge General English Online Test. The second component contains both (1) language proficiency Likert scale estimates of the two receptive (listening and reading) and two productive (speaking and writing) skills³, and (2) pseudo ‘situations’ that are based on the descriptor scales of the Common European Framework of Reference (CEFR). There are two sorts of filler items, where participants will indicate, by means of a Likert scale, (1) how they perceive each test charts their vocabulary, and (2) whether they appreciate the English language.

The questionnaire starts out with a brief demographic component, where the participants address their age, gender, and native language. The DIALANG Vocabulary Test and the Cambridge General English Online Test in the language testing component, following the brief demographic chart, are not mentioned by name, but rather by the name ‘part 1’ and ‘part 2’ (translated from Dutch: part 1 signifies ‘deel 1’ and part 2 signifies ‘deel 2’). Accordingly, the participants are not aware that they are filling out actual language tests. In Appendix 9.1, the entire questionnaire is enclosed. It can be seen that part 1 consists of the DIALANG Vocabulary Test, and part 2 consists of the Cambridge General English Online Test. The DIALANG test items, that are employed in the current questionnaire, are based on the test items that subjects get presented with when they do not fill out the placement test. Usually, when test takers take the DIALANG Vocabulary Test, they should fill out the placement test, that estimates the test taker’s vocabulary size, and determines which test items will be presented in the test. In the placement test, test takers will be presented with a collection of words, of which they have to

³ Inspiration for these scales was drawn from Klatter & Weltens (2017: 9). This study contains a self-assessment component as well, in which subjects indicated how they estimate their own language skills in general on a five-point scale. The scale ranges from 1 (very poor) to 5 (very good).

determine whether these are real words or not, by respectively clicking on the ‘Yes’ or ‘No’ button.

The DIALANG Vocabulary Test contains 30 questions, and the Cambridge General English Online Test contains 25 questions. The correct answer solutions of each question can be found in the answer grid in Table 2 in Appendix 9.2. The item types of both tests can be charted as follows:

Table 1. Outline of all the Item Types in the Language Testing Component				
			Items DIALANG Vocabulary Test (n = 30)	Items Cambridge General English Online Test (n = 25)
Item type	Multiple choice	Conventional	(n = 6)	(n = 25)
		Drop-down menu	(n = 9)	
	Gap-filling		(n = 5)	
	Short answer		(n = 10)	

The two different tests in the language testing component in the current questionnaire could be classified as a diagnostic test and a proficiency test. On the DIALANG website⁴, it is stated that the initial DIALANG project was conducted under the SOCRATES program ‘LINGUA 2’, with the assistance of the Commission of the European Communities. DIALANG could be classified as a diagnostic test, that offers the testing of five modules (reading, writing, listening, grammatical structures and vocabulary) in fourteen languages (Hughes, 1989: 16). In general, diagnostic tests are primarily designed to diagnose particular linguistic facets. According to Benmostefa (2008), the objective of diagnostic language tests is three-fold: (1) to offer learners a way to continue learning with their own personal learning plan (i.e. learning path), (2) to provide a way for learners to test their language skills, and (3) to offer information to the learners about their strengths and weaknesses (Benmostefa, 2008: 4). According to Hughes (1989), diagnostic tests are mainly intended to assess what still needs to be learnt. This is fairly clear at the level of the broad language skills (Hughes, 1989: 15).

⁴ For further reference, see: <https://dialangweb.lancaster.ac.uk>.

On the Cambridge English website⁵, there are four online English tests: (1) General English, (2) English Schools, (3) English Business, and (4) Young Learners. A test where the general command of English is assessed (e.g. the Cambridge General English Online test) could then accordingly be considered as a proficiency test. According to Hughes (1989), proficiency tests are designed to assess language ability, no matter what experience the test taker may have had in that language. Therefore, the content of a proficiency test is not dependent on the content or aims of a language course that could have been taken by people taking the test. Instead, it is based on a specification of what candidates in the language should be able to do in order to be considered proficient (Hughes, 1989: 11). On the Cambridge English website it is stated that the ‘Test Your English’ application is not a Cambridge English exam, and the scores and language levels are rather approximate. The score on this exam cannot be used as evidence of a valid formal language qualification.

As previously mentioned, after having completed the language testing component, the participants fill out the self-assessment component. The first part of the self-assessment component consists of five-point Likert scales where the subjects can mark their levels of (1) listening comprehension, (2) oral proficiency, (3) reading ability, and (4) writing ability. The second part of the self-assessment component is based on the level descriptions in the document ‘overzicht van alle tekstkenmerken’ (translated in English: overview of all text attributes) on the Dutch website⁶ of the Common European Framework of Reference. This part consists of three questions, where the levels A1 up to and including C2 represent situations. The situations are paraphrased in more straightforward language use. In the third question, there is no description for level A1, therefore this question only contains the descriptions for the levels A2 up to and including C2. The first question is based on the conversation component (receptive textual features, vocabulary) on page five in the overview document of all text attributes. In the questionnaire, this description is called ‘Praten in het Engels: kwaliteit van de woordenschat’ (translated in English: Talking in English: Quality of the vocabulary). The second question is based on the conversation component (productive textual features, range of vocabulary) on page seven in the overview document of all text attributes. In the questionnaire, this description is called ‘Bereik van de woordenschat’ (translated in English: Scope of the vocabulary). The third question is based on the writing component (productive textual features, vocabulary command)

⁵ For further reference, see: <https://www.cambridgeenglish.org/test-your-english/>.

⁶ For further reference, see: <http://www.erk.nl/docent/niveaubeschrijvingen/>.

on page twenty-two in the overview document of all text attributes. In the questionnaire, this description is called ‘Beheersing van de woordenschat’ (translated in English: Vocabulary command). The reason not all descriptions of the overview document of all text attributes are used, is that most of the descriptions are identical. Therefore, it has been chosen to solely integrate the aforementioned descriptions.

4.3 Procedure

Primarily, it is to be clarified that the participants fill out the questionnaire on their own terms. This implies that they can partake in the experiment in any setting (e.g. at home). It is intended that the questionnaire will not take longer than twenty minutes, hence the total duration of the task will be twenty minutes likewise. In the preface of the questionnaire, it is mentioned that the questionnaire is part of an investigation of the English vocabulary of various population groups (i.e. in Dutch society). Furthermore, it is clarified that the participants can only complete the survey once, and that the survey will close on April 17, 2020. If the participants have any questions with regard to the questionnaire, they can send an e-mail to the author of the current thesis.

The questionnaire commences with collecting essential demographic data (i.e. age, gender, and native language). Subsequently, the first part of the language testing component will follow: the DIALANG Vocabulary Test. After this first part, the participants fill out a five-point Likert scale that addresses whether they think the previous exercises accurately chart their English vocabulary knowledge. Hereafter, the subjects start with the second part of the language testing component: the Cambridge General English Online Test. Likewise, the participants fill out the five-point Likert scale that addresses whether they think the previous exercises accurately chart their English vocabulary knowledge. After the language testing component, there will be a component – named ‘Knowledge of English’ (translated in Dutch: Kennis van het Engels) – that addresses part of the demographic data of the participant (i.e. educational background), an attitudinal question regarding the appreciation of the English language, and the questions regarding self-assessment of the participants’ English vocabulary (see Chapter 9.1 for the entire questionnaire). It is to be clarified that all questions are mandatory, and the participants can solely submit their questionnaire when they have filled in all questions. After having filled out the entire questionnaire, the participants are being thanked for their participation in the experiment.

4.4 Design

The current research establishes a **correlational research design**, which could be classified under the quantitative research design category. The variables are (1) the CEFR levels on both the (1a) DIALANG Vocabulary Test, and the (1b) Cambridge General English Online Test; and (2) the self-assessment reports of both the (2a) self-assessment Likert-scale ratings concerning listening comprehension, oral proficiency, reading ability and writing ability, and (2b) actual CEFR descriptors of vocabulary knowledge. In correlational research, it is attempted to determine the extent of a relationship between two or more variables by means of reporting on statistical data. In the correlational design type, the links between and among a number of factors are investigated and interpreted. This research type could identify trends and patterns in data, but it does not establish causes for certain observed patterns in its analysis. This implies that the basis of the current research is not to observe cause and effect, and no variables are manipulated. The variables are only identified and observed as occurring in a natural environment.

4.5 Qualitative and Statistical Analyses

Primarily, it is to be noted that the three aforementioned sampling methods (convenience sampling, voluntary response sampling, and snowball sampling) are commonly accepted as the data being obtained from the participants rather being exploratory, where an initial understanding of specific outcomes obtained by the participants of the target population is achieved. Therefore, the data in the results section do not necessarily relate to the entire target population, i.e. all members of Dutch society.

The analysis of the results starts with section 5.1 where all the answers given to the two language tests are substantively studied. The importance of this lies in the recognition of certain patterns, such as many participants providing the false answer to particular questions, which could implicate that the question is too difficult. It could also be demonstrated that distractors of certain questions are being chosen over the correct answer. This section provides the reader with tables to present the precise summary of all questions. Furthermore, this section outlines the selected answers to the self-assessment items as well. After the qualitative analysis of the provided answers, the statistical analyses will be performed in sections 5.2 and 5.3. With regard to the statistical processing of the data obtained in the experiment, the reliability of all employed

scales, Cronbach's Alpha, will primarily be calculated. Statistical analyses that will be used to analyze and report on any correlational links between CEFR levels of both tests and self-assessment reports are Pearson product-moment correlation coefficients. In Blanche and Merino (1989), it is argued that regarding Pearson product-moment correlation coefficients, "[...] values ranging from .50 to .60 are common, and higher ones not uncommon. What this means is that a set of self-assessments (such as answers to a questionnaire) tends to carry about the same weight as any of the various parts (subtests) of a standardized testing instrument [...]" (Blanche & Merino, 1989: 315-324). Therefore, correlations of .50 and higher are considered as a significant correlation between the two variables (1) the CEFR levels on both the (1a) DIALANG Vocabulary Test, and the (1b) Cambridge General English Online Test; and (2) the self-assessment reports of both the (2a) self-assessment Likert-scale ratings concerning listening comprehension, speaking, reading and writing skills, and (2b) actual CEFR descriptors of vocabulary knowledge. Furthermore, to report on the extent to which the self-assessment items are able to predict actual CEFR levels on both tests, multiple regression analyses will be performed. The multiple regression analyses will be employed to test for differences between the two different language tests, and the demographic discrepancies in the third sub-question of the current research. With the multiple regression analyses, it is aimed to test if there is a set of predictors that best predict the outcome measure (i.e. the CEFR levels on the language testing component), meaning these predictors need not be causally related, only correlated. For all predictors, the multiple regression analysis generally assumes that if the p-value of the analysis is $p \leq .05$, there is statistical significance. However, in order to determine that a predictor is wrongly excluded from the equation, values ranging between $.05 \leq p \leq .10$ are seen as marginally statistically significant. The self-assessment items that fall within the scope of these p-values could therefore be included in the existence of unique variance. After all, to wrongfully select a predictor as having unique variance could be considered less harmful than wrongfully excluding that predictor.

5. Results

The first section of this chapter consists of a report on the frequency of all selected answer options on both tests of the language testing component, and the self-assessment component. This analysis will be performed by means of analyzing, and reporting on the Google Forms frequency output in a rather qualitative manner. For the selected answers on the self-assessment component, it will be reported on the SPSS frequency output regarding these items. The

quantitative analyses in the two subsequent sections (sections 5.2 and 5.3) were undertaken with SPSS.

5.1 Distribution of the Selected Answer Options in the Language Testing Component and the Self-Assessment Component

In this section, the given answers to the multiple choice items are outlined per each item, on the grounds that it gives an idea of how the current sample of the population took the tests. From that, we can deduce that for instance certain answer options are not chosen at all, or that other answer options are preferred over the correct answer option. In the following two subsections, the reader will be provided with a report on the answers to respectively the DIALANG Vocabulary Test – including the answers to the multiple choice questions, and the open short answer questions – and the answers to the Cambridge General English Online Test.

5.1.1 Report of the Provided Answers to the DIALANG Vocabulary Test

In Table 2, the distribution of all selected answers to the multiple choice items of the DIALANG Vocabulary Test is provided, including the frequency with which they occur and the relative frequency in relation to the other answer options. These items include items 1 up to and including 7, 17 up to and including 21, and items 26 and 29.

Table 2. Distribution of the Selected Multiple Choice Answer Options of the DIALANG Vocabulary Test

Item	Answer Option	Frequency (n)	Relative Frequency (%)
1. Kies het woord dat het best in _____ past: It may be possible to _____ damages against a local authority for not taking care of the roads well enough.	claim	63	86.3
	sue	9	12.3
	bet	1	1.4
2. Welk woord kan NIET worden toegevoegd aan het	why	71	97.3
	how	1	1.4
	one	1	1.4

woord 'any' om een nieuw
woord te maken?

3. Kies het woord dat het best	compete	59	80.8
in _____ past: I can't	oppose	7	9.6
_____ with your offer.	struggle	4	5.5
Anyone would take pizza	examine	3	4.1
instead of a soup mix.			
4. Kies het woord dat het best	iron	50	68.5
in _____ past: Don't wait	steel	12	16.4
any longer, you have to strike	metal	11	15.1
while the _____ is hot.			
5. Kies het woord dat het best	deputy	70	95.9
in _____ past: The armed	shelter	2	2.7
thief shot the sheriff, and	fortune	1	1.4
injured the _____ standing			
next to him.			
6. Kies het woord dat het best	bread	59	80.8
in _____ past: The	oil	9	12.3
_____ and butter of my	salt	5	6.8
life? I don't know ... The			
family? And my work.			
Making a good living, I			
suppose.			
7. Kies het woord dat	regard	54	74
ongeveer hetzelfde betekent	expect	11	15.1
als het woord 'consider'.	relate	7	9.6
	promise	1	1.4

17. Kies het woord dat hetzelfde betekent als het woord in HOOFDLETTERS in de volgende zin: There are a number of books and videos on the market, but it's still hard to learn 'tai chi' without personal TEACHING.	instruction	71	97.3
	adaption	2	2.7
18. Kies het woord dat hetzelfde betekent als 'motion'.	movement	70	95.9
	reacting	3	4.1
19. Kies het woord dat hetzelfde betekent als het woord in HOOFDLETTERS in de volgende zin: It has been my most sincere WISH for some time now.	desire	71	97.3
	request	2	2.7
20. Kies het woord dat het best op de puntjes (...) in beide zinnen past: We built a sandcastle. It ... when the waves came. Our plans ... , when the time allowed for completion was changed.	collapsed	58	79.5
	ruined	13	17.8
	devastated	2	2.7
21. Kies het woord dat het best in _____ past: The _____ of this factory is increasing.	output	60	82.2
	product	11	15.1
	aim	2	2.7

26. Kies het woord / de woorden dat/die hetzelfde betekent/beteken als het woord in HOOFDLETTERS in de volgende zin: The hotel has a small pleasant LOUNGE and bar, two terraces (one on the roof), a solarium (payable locally) and a sauna (free).	public room hallway restaurant	58 11 4	79.5 15.1 5.5
29. Kies het woord dat het tegenovergestelde van 'talkative' betekent.	quiet audible loud	65 5 3	89 6.8 4.1

As can be seen in Table 2, some items have been answered with relatively unanimous answers. There are no items where all participants selected the correct answer option. However, there are also no items where one or more distractors are selected over the correct answer. The items where participants selected relatively unanimous answers are item numbers 2, 5, 17, 18 and 19 with (more than) 70 participants out of 73 choosing the correct answer option. According to DIALANG, item 2 is a 'word formation' item, and item 5 is a 'word combination' item. Items 17, 18, and 19 are 'semantic relations' items. For items 17, 18, and 19, one distractor in addition to the correct answer option was selected; item 17 (correct answer: $n = 71$, and distractor: $n = 2$), item 18 (correct answer: $n = 70$, and distractor: $n = 3$) and item 19 (correct answer: $n = 71$, and distractor: $n = 2$). For item number 4, the highest number of distractors ($n = 23$) was chosen. This implies that for this item, the fewest number of participants filled out the correct answer option. According to DIALANG, item 4 is a 'semantic relations' item. This item consists of a regular and fixed saying in English, meaning that a large proportion of this group of Dutch participants is not familiar with the saying 'you have to strike while the iron is hot'. For items 1, 2, 4, 5, 6, 20, 21, 26, and 29, two distractors in addition to the correct answer option were selected; item 1 (correct answer: $n = 63$, and distractors: $n = 10$), item 2 (correct answer: $n = 71$, and distractors: $n = 2$), item 4 (correct answer: $n = 50$, and distractors: $n = 23$), item 5 (correct answer: $n = 70$, and distractors: $n = 3$), item 6 (correct answer: $n = 59$, and distractors:

n = 14), item 20 (correct answer: n = 58, and distractors: n = 15), item 21 (correct answer: n = 60, and distractors: n = 13), item 26 (correct answer: n = 58, and distractors: n = 15), and item 29 (correct answer: n = 65, and distractors: n = 8). For items 3 and 7, three distractors in addition to the correct answer option were selected; item 3 (correct answer: n = 59, and distractors: n = 14), and item 7 (correct answer: n = 54, and distractors: n = 19). Item 7 does not consist of a regular saying of the English language, but it is a ‘semantic relations’ item as well. A relatively high number of participants chose the distractor (n = 19), meaning that the participants might not be familiar with the semantic meaning of the word ‘consider’. It should be mentioned that, overall, there does not seem to be a connection to item type according to DIALANG and the size of the number of participants that chose the distractor instead of the correct answer option.

The DIALANG Vocabulary Test contains ample open questions as well. These questions are short answer questions, where participants should finish a given word, or write a short answer in the gap. In the following section, it will be outlined which answers the participants filled in. In Table 3a-p, all answers are provided. It should be mentioned that the relative frequency is rounded to one decimal place, which implies that the total percentage may not be 100 in some cases. However, it has been decided to provide the reader with the relative frequency after all, in order to provide a clear overview of the relative distribution of the selected answer options.

Table 3a. Outline of the Answers to Item 8 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
8. Schrijf het	ment	62	84.9
ontbrekende deel	ing	2	2.7
van het woord met	ped	1	1.4
het balkje hieronder	ping	1	1.4
op: The future	er	1	1.4
develop_____ of	ments	1	1.4
the European Union	depend	1	1.4
depends on what the	ed	1	1.4
member states want	state	1	1.4
out of Europe.	?	1	1.4
	more	1	1.4

Item 8, overall, has been filled out quite accurately with the largest proportion of the participants ($n = 62$) filling out the correct answer ‘ment’. Nine participants each filled the blank with a different answer. Two participants filled in ‘ing’, which makes a valid word as well, but is not correct in this context. Of the nine participants that each filled out a different answer, three of these answers (‘er’, ‘ments’, and ‘ed’) result in a legitimate word. Five of these answers (‘ped’, ‘ping’, ‘depend’, ‘state’ and ‘more’) result in a pseudoword. One participant did not know the answer to this item, and filled out a question mark, as all items are mandatory.

Table 3b. Outline of the Answers to Item 9 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
9. Welk woord past het best in het balkje in de volgende woordenlijst? Good - better - best _____ - worse - worst	bad very wors okay best	69 1 1 1 1	94.6 1.4 1.4 1.4 1.4

Similar to item 8, for item 9, overall, the largest proportion of the participants ($n = 69$) filled in the correct answer ‘bad’. Four participants each filled the blank with a different answer. Three participants filled in valid words (‘very’, ‘okay’, ‘best’), which are not correct in this context after all. One participant filled in ‘wors’ which is a pseudoword. It is likely that this particular participant intended to write ‘worse’, which is not correct in the context as well.

Table 3c. Outline of the Answers to Item 10 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
10. Schrijf het woord op dat het best in _____ past. Het woord begint met de	victim victum victem victom	65 2 2 2	89.1 2.7 2.7 1.4

letter 'v': He is a	vomit	1	1.4
pathological killer.	victume	1	2.7
His first _____			
was 30-year-old			
Tamara Lind, a			
former girlfriend.			

For item 9, most participants (n = 65) filled in the correct answer ‘victim’. Seven participants appear to have intended to fill in ‘victim’, but were not aware of how to accurately spell the correct answer. Therefore, a number of pseudowords were delivered (‘victum’, ‘victem’, ‘victom’ and ‘victume’). One participant filled in a valid word, which is not correct in this context (‘vomit’).

Table 3d. Outline of the Answers to Item 11 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
11. Schrijf het	ex	48	65.7
ontbrekende deel	sup	10	13.7
van het woord met	im	4	5.5
het balkje hieronder	sur	3	4.1
op: I will try to	su	2	1.4
_____press my	tell	1	2.7
feelings more	o	1	1.4
openly, but I'm not	de	1	1.4
sure I can.	op	1	1.4
	un	1	1.4
	-	1	1.4

For item 11, the smallest number of participants (n = 48) filled out the correct answer option. This entails that 25 participants filled in an incorrect answer. Twenty participants filled in an answer that leads to a valid word (‘sup’, ‘im’, ‘sur’, ‘de’, ‘op’ and ‘un’), which, however, is not correct in this particular context. Two participants filled in ‘su’ which results in a pseudoword, similar to one participant that filled in ‘o’, resulting in a pseudoword as well. One participant

filled in ‘tell’, from which it could be stated that, in this context, it has a similar semantic meaning as ‘express’. One participant did not know the answer to this item, and filled in a dash, as all items are mandatory.

Table 3e. Outline of the Answers to Item 12 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
12. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: My latest novel was _____jected by the first three publishers, but with the fourth one I got lucky!	re	67	91.8
	in	3	4.1
	op	1	1.4
	r	1	1.4
	-	1	1.4

For item 12, most participants ($n = 67$) filled in the correct answer ‘re’. Three participants filled in ‘in’, which results in a valid word, but it is not accurate in this context. One participant appeared to have intended to fill in the correct answer, but probably made an error by eliminating the letter ‘e’. One participant filled in ‘op’, resulting in a pseudoword. It is likely that this participant intended to fill in ‘ob’, as this is a valid word after all. One participant did not know the answer to this item, and filled in a dash, as all items are mandatory.

Table 3f. Outline of the Answers to Item of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
13. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: My	er	64	87.7
	ee	5	6.9
	ees	1	1.4
	ment	1	1.4
	chief	1	1.4

employ_____	-	1	1.4
gave me a bonus for working overtime.			

For item 13, most participants (n = 64) filled in the correct answer ‘er’. Seven participants filled in answers resulting in legitimate words (‘ee’, ‘ees’, ‘ment’), which, however, are not correct in this context. One participant filled in ‘chief’ which relates to the word ‘employer’, which was intended to complete by filling in the gap. One participant did not know the answer to this item, and filled in a dash, as all items are mandatory.

Table 3g. Outline of the Answers to Item 14 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
14. Schrijf het woord	color	41	56.1
op dat het best in	colour	28	38.4
_____ past. Het	center	3	4.1
woord begint met de	cillary	1	1.4
letter 'c': The			
_____ of her eyes			
is brown.			

As can be seen in Table 3g, most participants (n = 41) chose the preferred spelling in American English (‘color’) for item 14, and 28 participants chose the word ‘colour’, which is preferred in all other main varieties of English. Three participants filled in ‘center’, which is a valid word, but not accurate in the current context. One participant filled in ‘cillary’ – a pseudoword. It is likely that this participant intended to fill in ‘ciliary’, as this is a part of the eye consisting of muscle tissue.

Table 3h. Outline of the Answers to Item 15 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
15. Beantwoord de	fence	64	87.7
vraag met één	fense	2	2.7

woord. Het woord	fench	2	2.7
begint met 'f': What	fens	1	1.4
is usually put around	fences	1	1.4
a garden to separate	flower	1	1.4
one house from	f	1	1.4
another?	-	1	1.4

For item 15, most participants (n = 64) filled in the correct answer ‘fence’. Six participants appear to have intended to fill in the correct answer, but were not aware of the accurate spelling (‘fense’, ‘fench’, ‘fens’ and ‘fences’). One participant filled in ‘f’, which was already provided in the inquiry of the item, meaning that this participant did not know the answer to this item, and filled in the minimum entity, as all items are mandatory. One further participant appeared to have not known the correct answer, and was therefore obliged to fill in a dash. One participant filled in ‘flower’ which is a legitimate word, but rather not in the present context.

Table 3i. Outline of the Answers to Item 16 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
16. Welk woord betekent ongeveer hetzelfde als het woord 'sick'? Dit woord begint met de letter 'i'.	ill	73	100

For item 16, all participants filled in the correct answer solution ‘ill’. All participants were therefore aware of the semantic relation of the words ‘sick’ and ‘ill’, and were aware that the word ‘ill’ was intended to elicit.

Table 3j. Outline of the Answers to Item 22 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
------	----------------	---------------	------------------------

22. Welk woord betekent het tegenovergestelde van het woord dat in de volgende zin in HOOFDLETTERS is geschreven? Schrijf dat woord, dat in _____ hoort, hieronder op. Het woord begint met een 'o': On a sunny day I usually go _____ to get some fresh air. Who wants to stay INSIDE anyway?	outside out on	62 10 1	84.9 13.7 1.4
--	--	---	---

From the answer grid in Appendix 9.2, it can be deduced that for item 22, both the answers ‘out’ and ‘outside’ are correct. Therefore, 72 participants filled in the correct answer, and only one participant filled in a false answer (‘on’). The combination of ‘go’ and ‘on’ is valid, but it does not mean the opposite of the word in capital letters (‘inside’).

Table 3k. Outline of the Answers to Item 23 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
23. Welk woord betekent het tegenovergestelde van het woord 'known'? Gebruik dat woord voor _____	unknown strange new	69 2 1	94.5 2.7 1.4

in de volgende zin.	fomilair	1	1.4
Schrijf het woord			
hieronder op: That			
beach was _____			
to us, because we			
lived so far away			
from it.			

For item 23, most participants (n = 69) filled in the correct answer ‘unknown’. This ‘word formation’ item could be deduced from the inquiry, as the antonym of ‘known’ is requested. The participants are not provided with, for instance, the first letter(s). Therefore, it could be considered explicable that two participants filled in the word ‘strange’, which has a similar lexical meaning as ‘unknown’. The same goes for the word ‘new’ – filled in by one participant – which is related to the lexical item ‘unknown’ as well. One participant seems to have intended to fill in ‘familiar’, yet actually filled in the pseudoword ‘fomilair’. The word ‘familiar’ is actually similar to the word ‘known’, and therefore it is likely that this participant was not aware that the antonym for ‘known’ was meant, but rather a synonym.

Table 31. Outline of the Answers to Item 24 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
24. Schrijf het woord	uncomfortable	60	81.8
dat het best in	oncomfortabel	1	1.4
_____ past	uncleaned	1	1.4
hieronder op. Het	unconvenient	1	1.4
woord begint met	?	1	1.4
'unc...': The seats	unclaen	1	1.4
were rather _____	uncontrolled	1	1.4
, and it was not easy	uncomfortable	1	1.4
to remain sitting	unclosed	1	1.4
down all that time.	uncomplete	1	1.4
	unclear	1	1.4
	unchairable?	1	1.4

uncle	1	1.4
-	1	1.4

For item 24, a high number of diverse answers ($n = 13$), that are not in line with the correct answer solution, has been provided. This may imply that this question is rather difficult to complete accurately. However, the largest proportion of participants ($n = 60$) filled in the correct answer ('uncomfortable'). Two participants provided a similar answer, yet with an incorrect spelling ('oncomfortabel' and 'unconfortable'). The former of these two incorrectly spelled words is rather equal to Dutch spelling, and the latter contains a spelling mistake. Two provided answers ('uncleaned' and 'unclear') are valid words, and it could be stated that they are relatively plausible in this context as well. The words 'uncontrolled', 'unclosed', 'uncomplete', and 'uncle' are valid words as well, but not plausible in this context. One participant appears to have intended to fill in the word 'inconvenient', which would be plausible in this context as well, but the participant spelled this word as 'unconvenient'. Two participants did not provide a textual answer to this item, and filled in a dash and a question mark, as all items are mandatory. One participant filled in a pseudoword ('unchairable?'), with the question mark implying that the participant was not aware of the intended answer. As there are two provided answers that seem as valid as the correct answer option, by the judgement of the author of this thesis, it could be stated that it might not be reliable and valid to enclose this item in the DIALANG Vocabulary Test, or that it may be suggested to alter this item. This claim is supported by the highest number of pseudowords and non-textual answers, such as dashes and question marks, meaning that this item may be too difficult.

Table 3m. Outline of the Answers to Item 25 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
25. Kies het woord	dog	58	79.5
dat het best op de	dogs	13	17.8
puntjes (...) in beide	rats	1	1.4
zinnen past: It's	-	1	1.4
raining cats and ...s.			
Treat somebody like			
a			

For item 25, the largest proportion of the participants filled in the correct answer ‘dog’ (n = 58). Some participants (n = 13) filled in ‘dogs’ which would have been correct if the only inquiry was to finish the sentence ‘It’s raining cats and ...[]’, and it would not have been required to finish the sentence ‘Treat somebody like a’ The catch in this item is therefore to decipher that the singular word ‘dog’ is meant, as it fits in both sentences. One participant filled in ‘rats’, meaning that this participant was not familiar with both regular sayings. The remaining participant was not certain what they should fill in, and accordingly filled in a dash.

Table 3n. Outline of the Answers to Item 27 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
27. Schrijf het	piece	54	73.9
ontbrekende deel	?	3	4.1
van het woord met	s	3	4.1
het balkje hieronder	class	2	2.7
op: The Guggenheim	-	2	2.7
Museum is being	plan	2	2.7
hailed as the greatest	work	2	2.7
architectural	pièce	1	1.4
master_____ of	full	1	1.4
this century.	is	1	1.4
	y	1	1.4
	of	1	1.4

The outline of the answers to item 27 shows that the largest proportion of participant (n = 56) filled in the correct answer ‘piece’ or ‘work’. One participant filled in a similar lexical entity, but with a deviant spelling (‘pièce’). Eight of the provided answers (‘s’, ‘class’, ‘plan’, and ‘y’) result in a valid word, which, however, is not legitimate in this context. Three participants filled in answers that result in pseudowords (‘full’, ‘is’, and ‘of’). Five participants did not know the answer to this item, and filled in a dash and a question mark, as all items are mandatory.

Table 30. Outline of the Answers to Item 28 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
28. Schrijf het woord	ambulance	65	89
op dat het best in	?	3	4.1
_____ past. Het	-	2	2.7
woord begint met de	automobil	1	1.4
letter 'a': He was	ambulanc	1	1.4
badly injured and	ambulance	1	1.4
they took him to			
hospital in an			
_____.			

For item 28, the largest proportion of the participants filled in the correct answer ‘ambulance’ (n = 65). Two participants appear to have intended to fill in the correct answer, yet failed to provide the correct spelling (‘ambulanc’ and ‘ambulance’). One participant filled in a lexical item, which is not existent in the English language (‘automobil’). Five participants were not certain what the answer to this item was, and therefore filled in a dash or a question mark.

Table 3p. Outline of the answers to item 30 of the DIALANG Vocabulary Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
30. Welk woord	encourage	57	78.1
betekent het	incourage	7	9.6
tegenovergestelde	courage	5	6.8
van het woord dat in	stimulate	1	1.4
de volgende zin in	?	1	1.4
HOOFDLETTERS	support	1	1.4
is geschreven?	-	1	1.4
Schrijf dat woord			
hieronder op: But the			
state should go			
further to			

DISCOURAGE

impressionable

children from

smoking, says

political activist

Steven Brown.

Item 28 has been filled in with the largest proportion of the participants having filled in the correct answer ‘encourage’ (n = 57). Seven participants appear to have intended to fill in the correct answer, yet failed to provide the correct spelling (‘incourage’). Five participants filled in ‘courage’, which is a valid word, but it is not the antonym of ‘discourage’. One participant filled in ‘stimulate’ and one participant filled in ‘support’, which are words that are adverse to the word ‘discourage’, but they are not the intended and correct antonym. Two participants were not certain what the answer to this item was, and therefore filled in a dash or a question mark.

5.1.2 Report of the Provided Answers to the Cambridge General English Online Test

As can be seen in Appendix 9.1, the Cambridge General English Online Test solely consists of multiple choice answers. In Table 4, all the answer options that the participants selected are outlined.

Table 4. Distribution of the Selected Multiple Choice Answer Options of the Cambridge General English Online Test

Item	Answer Options	Frequency (n)	Relative Frequency (%)
1. Can I park here?	Only for half an hour.	73	100
2. What colour will you paint the children's bedroom?	We can't decide.	73	100

3. I can't understand this email.	Would you like some help?	68	93.2
	I suppose you can.	5	6.8
4. I'd like two tickets for tomorrow night.	I'll just check for you.	67	91.8
	How much did you pay?	4	5.5
	Afternoon and evening	2	2.7
5. Shall we go to the gym now?	I'm too tired.	68	93.2
	It's very good.	4	5.5
	Not at all.	1	1.4
6. His eyes were bad that he couldn't read the number plate of the car in front.	so	73	100
7. The company needs to decide and for all what its position is on this point.	once	61	83.6
	first	9	12.3
	here	2	2.7
	finally	1	1.4
8. Don't put your cup on the of the table – someone will knock it off.	edge	69	94.5
	outside	3	4.1
	border	1	1.4

9. I'm sorry - I didn't to disturb you.	mean suppose	72 1	98.6 1.4
10. The singer ended the concert her most popular song.	with by	70 3	95.9 4.1
11. Would you mind these plates a wipe before putting them in the cupboard?	giving getting doing	68 3 2	93.2 4.1 2.7
12. I was looking forward at the new restaurant, but it was closed.	to eat eating to eating	54 10 9	74 13.7 12.3
13. tired Melissa is when she gets home from work, she always makes time to say goodnight to the children.	No matter how	73	100
14. It was only ten days ago she started her new job.	since that after	45 27 1	61.6 37 1.4
15. The shop didn't have the shoes I	ordered	73	100

wanted, but they've
..... a pair specially
for me.

16. Have you got	about	66	90.4
time to discuss your	planned	7	9.6
work now or are you			
..... to leave?			

17. She came to live	almost	68	93.2
here a month	already	4	5.5
ago.	beyond	1	1.4

18. Once the plane is	unfasten	66	90.4
in the air, you can	unlock	6	8.2
..... your seat belts if	untie	1	1.4
you wish.			

19. I left my last job	opportunity	38	52.1
because I had no to travel.	possibility	33	45.2
	place	1	1.4
	position	1	1.4

20. It wasn't a bad	little	58	79.5
crash and damage was done to my car.	small	11	15.1
	light	3	4.1
	mere	1	1.4

21. I'd rather you	would explain	36	49.3
..... to her why we	explained	22	30.1
can't go.	will explain	10	13.7
	to explain	5	6.8

22. Before making a decision, the leader considered all of the argument.	perspectives	37	50.7
	sides	29	39.7
	features	7	9.6
23. This new printer is recommended as being reliable.	highly	65	89
	strongly	8	11
24. When I realised I had dropped my gloves, I decided to my steps.	retrace	55	75.3
	return	13	17.8
	resume	5	6.8
25. Anne's house is somewhere in the of the railway station.	vicinity	30	41.1
	district	21	28.8
	region	18	24.7
	quarter	4	5.5

In the current report of the distribution of selected answer options, it could be made up that four items have been filled in correctly by all participants (items 1, 2, 6, 13, and 15). For five items, a proportion of the participants chose one distractor, in addition to the correct answer; item 3 (correct answer: n = 68, and distractor: n = 5), item 9 (correct answer: n = 72, and distractor: n = 1), item 10 (correct answer: n = 70, and distractor: n = 3), item 16 (correct answer: n = 66, and distractor: n = 7), and item 23 (correct answer: n = 65, and distractor: n = 8). For seven items, a proportion of the participants chose two distractors, in addition to the correct answer; item 4 (correct answer: n = 67, and distractors: n = 6), item 5 (correct answer: n = 68, and distractors: n = 5), item 8 (correct answer: n = 69, and distractors: n = 4), item 11 (correct answer: n = 68, and distractors: n = 5), item 17 (correct answer: n = 68, and distractors: n = 5), item 18 (correct answer: n = 66, and distractors: n = 7), and item 24 (correct answer: n = 55, and distractors: n = 18). For three items, the distribution of answer options consists of three selected answer options as well, but for these items, the distractor is selected over the correct answer. This is the case for items 12, 14, and 22. For item 12, the correct answer is 'to eating' (n = 9), while most participants chose the distractor 'to eat' (n = 54), and few participants chose

the distractor ‘eating’ (n = 10). It appears that this sample of the Dutch population experiences difficulties in accurately using the present continuous, as required in item 12. For item 14, the correct answer is ‘that’ (n = 27). However, most participants chose the distractor ‘since’ (n = 45), and one participant chose the distractor ‘after’ (n = 1). For item 22, the correct answer is ‘sides’ (n = 29), and most participants chose the distractor ‘perspectives’ (n = 37), or the other distractor ‘features’ (n = 7). For three items, a proportion of the participants chose three distractors, in addition to the correct answer; item 7 (correct answer: n = 61, and distractors: n = 12), item 19 (correct answer: n = 38, and distractors: n = 35), and item 20 (correct answer: n = 58, and distractors: n = 15). For two items, the distribution of answer options consists of four selected answer options as well, but for these items, the distractor is selected over the correct answer. This is the case for items 21 and 25; for item 21, ‘explained’ is the correct answer, (n = 36), and the distractors are chosen 37 times in total. For item 25, the correct answer is ‘vicinity’ (n = 30), and the distractors are chosen 43 times.

Now that the outline of all selected answers in the language testing component has been provided, it is also sensible to provide the reader with the selected answer options on the self-assessment component. Therefore, in Table 5 below, the selected answer options of the self-assessment component are outlined, by means of providing the frequency output from SPSS. It can be seen that, overall, for the Likert scale ratings, the participants do not self-assess themselves as ‘very bad’ or ‘bad’ that often. For the CEFR descriptors, the self-assessed levels are fairly divergent. However, very low (A1) and very high (C2) levels are not occurring that often.

Table 5. Distribution of the Selected Answer Options of the Self-Assessment Component

Self-assessment Item	Frequency Output	
Likert scale ratings of listening comprehension	Very bad	0
	Bad	1
	Neutral	9
	Good	39
	Very good	24
Likert scale ratings of oral proficiency	Very bad	2
	Bad	7
	Neutral	22

	Good	31
	Very good	11
Likert scale ratings of reading ability	Very bad	0
	Bad	2
	Neutral	14
	Good	33
	Very good	24
Likert scale ratings of writing ability	Very bad	1
	Bad	13
	Neutral	26
	Good	24
	Very good	9
Quality of the vocabulary (oral proficiency)	A1	2
	A2	19
	B1	26
	B2	11
	C1	13
	C2	2
Scope of the vocabulary	A1	1
	A2	4
	B1	21
	B2	25
	C1	22
	C2	0
Vocabulary command (writing ability)	A1	0
	A2	2
	B1	12
	B2	29
	C1	25
	C2	5

5.2 Correlations between the post-test Self-Assessment Component and the CEFR

Levels of Vocabulary Knowledge

Before all else, the reliability of the self-assessment and the language testing scales was calculated. The internal consistency of the self-assessment Likert scale ratings (listening comprehension, oral proficiency, reading ability, and writing ability) is high ($\alpha = .842$). The Cronbach's Alpha indeed shows that it is possible to measure self-assessment with. Eliminating either one of the items results in a lower internal consistency. The reliability of the self-assessment items based on the CEFR is even higher ($\alpha = .858$). However, if the item 'vocabulary command (writing ability)' is not included, there will be a higher internal consistency between the scales resulting in a higher reliability of the scale ($\alpha = .875$). The reliability statistics of all the self-assessment items – including the three CEFR descriptors, and the four Likert scale ratings – show that all seven items contain very high internal consistency ($\alpha = .912$). This entails that the self-assessment component, as a construct, may be very well employed to measure self-assessment with in this particular study. Eliminating either one of these seven items will lead to a lower internal consistency. The internal consistency of both of the items, where all the reported CEFR levels according to the tests are recorded, is rather questionable ($\alpha = .603$). This means that it is equivocal to state that the CEFR levels of both the DIALANG Vocabulary Test and the Cambridge General English Online Test measure the construct of 'vocabulary knowledge' in a consistent manner. However, for the current thesis it is not intended for the entire language testing component to correlate with the self-assessment component as a whole. It is rather intended to plot either the DIALANG Vocabulary Test results, or the Cambridge General English Online Test results, both in terms of CEFR levels, against the self-assessment items, be it as a unified construct or as separate scales (both will be performed in sections 5.2.1 and 5.2.2). The correlations between the language testing component and the self-assessment component will be calculated in the following two sections.

5.2.1 Correlations between the post-test Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

Primarily, this section will provide the reader with correlations of each of the seven self-assessment items with the CEFR levels according to the DIALANG Vocabulary Test. Hereby, the discrepancies between either (a) CEFR descriptors of vocabulary knowledge, or (b) the Likert scale ratings concerning listening comprehension, oral proficiency, reading ability, and writing ability could be reported on.

Table 6a. Pearson Correlations of the four Likert Scale Ratings of the Self-Assessment Component and CEFR Levels according to the DIALANG Vocabulary Test

		CEFR levels according to the DIALANG Vocabulary Test
Likert scale ratings of listening comprehension	Pearson Correlation Sig. (2-tailed)	.406** .000
Likert scale ratings of oral proficiency	Pearson Correlation Sig. (2-tailed)	.310** .008
Likert scale ratings of reading ability	Pearson Correlation Sig. (2-tailed)	.288* .013
Likert scale ratings of writing ability	Pearson Correlation Sig. (2-tailed)	.402** .000

*. Correlation is significant at the .05 level (2-tailed).

**. Correlation is significant at the .01 level (2-tailed).

Table 6b. Pearson Correlations of the three CEFR Descriptors of the Self-Assessment Component and CEFR Levels according to the DIALANG Vocabulary Test

		CEFR levels according to the DIALANG Vocabulary Test
Quality of the vocabulary (oral proficiency)	Pearson Correlation Sig. (2-tailed)	.350** .002
Scope of the vocabulary	Pearson Correlation Sig. (2-tailed)	.387** .001
Vocabulary command (writing ability)	Pearson Correlation Sig. (2-tailed)	.423** .000

**. Correlation is significant at the .01 level (2-tailed).

Although all correlations are significant, overall, these correlations are rather low/weak ($.30 \leq r \leq .50$), apart from the correlation between the Likert scale ratings of reading ability, and the CEFR levels according to the DIALANG Vocabulary Test, which barely shows a correlation

($r = .288$). Correlations between 0.00 and .30 are considered as demonstrating little or no correlation. The correlation between ‘vocabulary command (writing ability)’ and the CEFR levels according to the DIALANG vocabulary test is the highest ($r = .423$). Nonetheless, these low correlations do not necessarily mean that there exists no correlation between self-assessment and actual CEFR levels one achieves on the language test. It is rather interesting to investigate whether the self-assessment component, as a construct, correlates with the test results in the language testing component. Therefore, in the following passage of this section, two multiple regression analyses will be carried out. The first analysis will test if, when considered separately, both (a) the Likert scale ratings, and (b) the CEFR descriptors of vocabulary knowledge, which are both part of the self-assessment component, are able to predict the CEFR levels the participants will achieve on the DIALANG Vocabulary Test. The second multiple regression analysis will test if the self-assessment component as a unified construct is able to predict the CEFR levels the participants will achieve on the DIALANG Vocabulary Test.

Firstly, the multiple regression analysis with only Likert scale ratings, and only CEFR descriptors as construct of self-assessment will be carried out. The first test will consist of assigning all the four Likert scale ratings of the self-assessment component as the independent variables, i.e. the predictors. The dependent variable consists of the CEFR levels according to the DIALANG Vocabulary Test.

Table 7. Model Summary of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.464	.215	.169	1.22031

The correlation coefficient R is .464, which implies that there exists a low or weak positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .169, which entails that the independent variables account for the CEFR level the participants may achieve on the DIALANG Vocabulary Test in 16.9% of the cases. From the ANOVA test of the multiple regression analysis, it could be concluded that the overall regression model is

statistically significant, $F(4, 68) = 4.65$, $p < .005$, $R^2 = .22$. According to the DIALANG Vocabulary Test, the independent variables can account for a significant amount of variance in the CEFR levels, which means that the four Likert scale self-assessment items can predict the DIALANG Vocabulary Test CEFR rates, combined as one category.

Table 8. ANOVA Results of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	27.723	4	6.931	4.654	.002
Residual	101.263	68	1.489		
Total	128.986	72			

From the coefficient results in the multiple regression analysis output, it could be concluded that the unique variance of each of the four independent variables, or predictors, is statistically significant for one out of four independent variables. This independent variable consists of the Likert scale ratings of listening comprehension ($p = .045$). This implies that this individual item uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. One variable, the Likert scale ratings of writing ability, could be considered as being marginally statistically significant ($p = .092$).

Table 9. Coefficients Results of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

Predictor	B	SE B	β	t	p-value
Likert scale ratings of listening comprehension	.566	.277	.293	2.039	.045
Likert scale ratings of oral proficiency	.048	.208	.034	.230	.819

Likert scale ratings of reading ability	-.133	.260	-.079	-.513	.610
Likert scale ratings of writing ability	.373	.218	.269	1.711	.092

Subsequently, the multiple regression analysis for the CEFR descriptors as a self-assessment instrument was carried out in order to report in what way these predictors account for the CEFR levels the participants will achieve on the DIALANG Vocabulary Test.

Table 10. Model Summary of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.456	.208	.173	1.21709

The correlation coefficient R is .456, which implies that there exists a low or weak positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .173, which entails that the independent variables are able to predict the CEFR level the participants may achieve on the DIALANG Vocabulary Test in 17.3% of the cases. This is slightly higher than the Likert scale ratings part of the self-assessment component is able to predict the test results of the DIALANG Vocabulary Test. From the ANOVA test of the multiple regression analysis, it could be concluded that the overall regression model is statistically significant, $F(3, 69) = 6.03, p < .005, R^2 = .21$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test. Accordingly, it could be said that the CEFR descriptors, combined as one group, can predict the DIALANG Vocabulary Test CEFR levels.

Table 11. ANOVA Results of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	26.777	3	8.926	6.026	.001
Residual	102.209	69	1.481		
Total	128.986	72			

From the coefficient results in the multiple regression analysis output (see Table 11), it could be concluded that the unique variance of each of the three independent variables, or predictors, is statistically significant for one out of three independent variables. This independent variable is the ‘vocabulary command (writing ability)’ variable ($p = .035$). This implies that this individual item uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for.

Table 12. Coefficients Results of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

Predictor	B	SE B	β	t	p-value
Quality of the vocabulary (oral proficiency)	-.020	.210	-.018	-.096	.924
Scope of the vocabulary	.307	.251	.221	1.225	.225
Vocabulary command (writing ability)	.445	.207	.304	2.149	.035

In the following passage of this section, a multiple regression analysis will be carried out in order to test if the self-assessment component as a unified construct is able to predict the CEFR levels the participants will obtain on the DIALANG Vocabulary Test. The independent variables, i.e. the predictors, are all the seven items that are employed for the self-assessment

component, and the dependent variable consists of the CEFR levels according to the DIALANG Vocabulary Test.

Table 13. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.515	.265	.186	1.20746

It can be deduced from Table 12, that the correlation coefficient R is .515, which implies that there exists moderate correlation between the independent variables and the dependent variable. *Adjusted R²* is .186, which entails that the independent variables account for the CEFR level the participants may achieve on the DIALANG Vocabulary Test in 18.6% of the cases. From the ANOVA test of the multiple regression analysis, it could be concluded that the overall regression model is statistically significant, $F(7, 65) = 3.35, p < .005, R^2 = .27$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test. It could accordingly be stated, that all self-assessment items, combined as one group, are able to predict the CEFR levels of the DIALANG Vocabulary Test.

Table 14. ANOVA Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	34.220	7	4.889	3.353	.004
Residual	94.767	65	1.458		
Total	128.986	72			

From the coefficient results in the multiple regression analysis output (see Table 14), it could be concluded that the unique variance of each of the seven independent variables, or predictors, is not statistically significant. This means that neither of these seven individual items uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. This means that the entire self-assessment component is

correlated internally to such a degree that when viewed separately, none of the independent variables provides any substantial amount of specific variation in predicting the dependent variables.

Table 15. Coefficients Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test

Predictor	B	SE B	β	t	p-value
Likert scale ratings of listening comprehension	.469	.282	.243	1.667	.100
Likert scale ratings of oral proficiency	-.192	.243	-.137	-.788	.433
Likert scale ratings of reading ability	-.330	.283	-.196	-1.166	.248
Likert scale ratings of writing ability	.256	.229	.185	1.118	.268
Quality of the vocabulary (oral proficiency)	.010	.220	.009	.046	.963
Scope of the vocabulary	.273	.277	.196	.983	.329
Vocabulary command (writing ability)	.388	.254	.265	1.531	.131

5.2.2 Correlations between the post-test Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

Primarily, this section will provide the reader with the correlations of each of the seven self-assessment items with the CEFR levels according to the Cambridge General English Online Test. Hereby, the discrepancies between either (a) the CEFR descriptors of vocabulary knowledge, or (b) the Likert scale ratings concerning listening comprehension, oral proficiency, reading ability, and writing ability could be reported on.

Table 16a. Pearson Correlations of the four Likert Scale Ratings of the Self-Assessment Component and CEFR Levels according to the Cambridge General English Online Test

		CEFR levels according to the Cambridge General English Online Test
Likert scale ratings of	Pearson Correlation	.452**
listening comprehension	Sig. (2-tailed)	.000
Likert scale ratings of oral	Pearson Correlation	.450**
proficiency	Sig. (2-tailed)	.000
Likert scale ratings of	Pearson Correlation	.440**
reading ability	Sig. (2-tailed)	.000
Likert scale ratings of	Pearson Correlation	.513**
writing ability	Sig. (2-tailed)	.000

**. Correlation is significant at the .01 level (2-tailed).

Table 16b. Pearson Correlations of the three CEFR descriptors of the Self-Assessment Component and CEFR Levels according to the Cambridge General English Online Test

		CEFR levels according to the Cambridge General English Online Test
Quality of the vocabulary	Pearson Correlation	.585**
(oral proficiency)	Sig. (2-tailed)	.000
Scope of the vocabulary	Pearson Correlation	.562**
	Sig. (2-tailed)	.000
	Pearson Correlation	.466**

Vocabulary command (writing ability)	Sig. (2-tailed)	.000
---	-----------------	------

** . Correlation is significant at the .01 level (2-tailed).

Some of the correlations between the seven self-assessment items and the CEFR levels according to the Cambridge General English Online Test are fairly higher than the correlations between the self-assessment items and the DIALANG Vocabulary Test. The Likert scale ratings of writing ability ($r = .513$), quality of the vocabulary (oral proficiency) ($r = .585$), and scope of the vocabulary ($r = .562$) demonstrate moderate positive correlation with the CEFR levels according to the Cambridge General English Online Test ($.50 \leq r \leq .70$). The other self-assessment items correlate fairly higher with the Cambridge General English Online Test than the self-assessment items do with the DIALANG Vocabulary Test. However, they still fall within the scope of $.30 \leq r \leq .50$, meaning the correlation is low/weak. Similar to section 5.2.1, this section will investigate whether the self-assessment component, as a construct, correlates with the test results of the Cambridge General English Online Test. Therefore, in the following passage of this section, two multiple regression analyses will be carried out as well. The first analysis will test if, examined individually, both (a) the Likert scale ratings, and (b) the CEFR descriptors of vocabulary knowledge, which are both part of the self-assessment component, are capable of predicting the CEFR levels the participants will achieve on the Cambridge General English Online Test. The second multiple regression analysis will test if the self-assessment component as a unified construct is able to predict the CEFR levels the participants will achieve on the Cambridge General English Online Test.

Firstly, the multiple regression analysis with only the Likert scale ratings as a construct, and only the CEFR descriptors as construct will be carried out. The first test will consist of assigning all the four Likert scale ratings of the self-assessment component as the independent variables, i.e. the predictors. The dependent variable consists of the CEFR levels according to the Cambridge General English Online Test.

Table 17. Model Summary of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.566	.321	.281	1.68110

The correlation coefficient R is .566, which implies that there exists a moderate positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .281, which entails that the independent variables account for the CEFR level the participants may achieve on the Cambridge General English Online Test in 28.1% of the cases. This is fairly higher than the DIALANG Vocabulary Test, for which the Likert scale ratings account for the CEFR level for 16.9%. From the ANOVA test of the multiple regression analysis, it could be concluded that the overall regression model is statistically significant, $F(4, 68) = 8.03$, $p < .0001$, $R^2 = .32$. According to the Cambridge General English Online Test, the independent variables can account for a significant amount of variation in the CEFR rates. It could therefore be stated that the four self-assessment items of the Likert scale, combined as one group, can predict the CEFR levels of the Cambridge General English Online Test.

Table 18. ANOVA Results of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	90.811	4	22.703	8.033	.000
Residual	192.175	68	2.826		
Total	282.986	72			

From the coefficient results in the multiple regression analysis output (see Table 18), it could be concluded that the unique variance of each of the four independent variables, or predictors, is not statistically significant. This means that neither of these four individual items uniquely accounts for variance in the CEFR levels according to the Cambridge General English Online Test, that the other items do not account for. This implies that the independent variables, i.e. the Likert scale ratings part of the self-assessment component, are correlated with each other to such a degree that none of the independent variables, when considered individually, offers a

significant amount of unique variance in predicting the dependent variable. However, the Likert scale ratings of writing ability variable is marginally statistically significant ($p = .064$).

Table 19. Coefficients Results of the Multiple Regression Analysis of the Likert Scale Ratings of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

Predictor	B	SE B	β	t	p-value
Likert scale ratings of listening comprehension	.535	.382	.187	1.401	.166
Likert scale ratings of oral proficiency	.276	.286	.133	.967	.337
Likert scale ratings of reading ability	.198	.358	.080	.554	.582
Likert scale ratings of writing ability	.566	.300	.275	1.884	.064

Subsequently, the multiple regression analysis for the CEFR descriptors as a self-assessment instrument was carried out in order to report in what way these predictors account for the CEFR levels the participants will achieve on the Cambridge General English Online Test.

Table 20. Model Summary of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.614	.377	.350	1.59890

The correlation coefficient R is .614, which implies that there exists a moderate positive correlation between the independent variables and the dependent variable. When comparing this correlation to the correlation of the CEFR descriptors in the self-assessment component and the CEFR levels according to the DIALANG Vocabulary Test, this correlation is relatively higher. *Adjusted R^2* is .350, which entails that the independent variables account for the CEFR level the participants may achieve on the Cambridge General English Online Test in 35% of the cases. This is slightly higher than the Likert scale ratings part of the self-assessment component is able to predict the test results of the DIALANG Vocabulary Test. From the ANOVA test of the multiple regression analysis, it could be concluded that the overall regression model is statistically significant, $F(3, 69) = 13.90, p < .0001, R^2 = .38$. According to the Cambridge General English Online Test, the independent variables can account for a significant amount of variation in the CEFR rates. Thus, it could be stated that the self-assessment items of the CEFR descriptors, combined as one group, can predict the CEFR levels of the Cambridge General English Online Test.

Table 21. ANOVA Results of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	106.589	3	35.530	13.898	.000
Residual	176.398	69	2.556		
Total	282.986	72			

From the coefficient results in the multiple regression analysis output (see Table 21), it could be concluded that the unique variance of each of the three independent variables, or predictors, is not statistically significant. This means that neither of these three individual CEFR descriptors uniquely accounts for variance in the CEFR levels according to the Cambridge General English Online Test, that the other items did not account for. This implies that the independent variables, i.e. the CEFR descriptors part of the self-assessment component, are correlated with each other to such a degree that none of the independent variables, when considered individually, offers a significant amount of unique variance in predicting the

dependent variable. However, the quality of the vocabulary variable (oral proficiency) is marginally statistically significant ($p = .056$).

Table 22. Coefficients Results of the Multiple Regression Analysis of the CEFR Descriptors of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

Predictor	B	SE B	β	t	p-value
Quality of the vocabulary (oral proficiency)	.537	.276	.323	1.944	.056
Scope of the vocabulary	.477	.329	.231	1.448	.152
Vocabulary command (writing ability)	.268	.272	.123	.984	.329

In the following passage of this section, a multiple regression analysis will be carried out in order to test if the self-assessment component as a unified construct is able to predict the CEFR levels the participants will achieve on the Cambridge General English Online Test. The independent variables, i.e. the predictors, are all the seven items that are employed for the self-assessment component, and the dependent variable consists of the CEFR levels according to the Cambridge General English Online Test.

Table 23. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

R	R ²	Adjusted R ²	Std. Error of the Estimate
.650	.423	.361	1.58484

The correlation coefficient R is .650, which implies that there exists moderate correlation between the independent variables and the dependent variable, similar to the correlation

coefficient of the self-assessment component and the CEFR levels according to the DIALANG Vocabulary Test. In this case, *Adjusted R²* is .361, which entails that the independent variables have the capacity of predicting the CEFR level the participants may achieve on the Cambridge General English Online Test in 36.1% of the cases. This is fairly higher than for the DIALANG Vocabulary Test (*Adjusted R²* = .186). The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant, $F(7, 65) = 6.81, p < .0001, R^2 = .42$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to Cambridge General English Online Test. It could accordingly be stated, that all self-assessment items, combined as one group, are able to predict the CEFR levels of the Cambridge General English Online Test.

Table 24. ANOVA Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

	Sum of Squares	df	Mean Square	F	Sig.
Regression	119.725	7	17.104	6.809	.000
Residual	163.262	65	2.512		
Total	282.986	72			

From the coefficient results in the multiple regression analysis output, it could be concluded that the unique variance of each of the seven independent variables, or predictors, is statistically significant for two out of seven independent variables. These items are the Likert scale ratings for writing ability ($p = .049$) and the CEFR descriptor of the quality of the vocabulary (oral proficiency) ($p = .046$). The latter variable relates to CEFR levels of oral proficiency and the former relates to Likert scale self-assessment estimates of writing ability. This implies that these individual items uniquely account for variance in the CEFR levels according to the Cambridge General English Online Test, that the other items did not account for.

Table 25. Coefficients Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test

Predictor	B	SE B	β	t	p-value
Likert scale ratings of	.258	.369	.090	.698	.488

listening					
comprehension					
Likert scale	-.217	.320	-.105	-.680	.499
ratings of oral					
proficiency					
Likert scale	-.145	.372	-.058	-.390	.698
ratings of					
reading ability					
Likert scale	.605	.301	.294	2.009	.049
ratings of					
writing ability					
Quality of the	.587	.288	.353	2.034	.046
vocabulary					
(oral					
proficiency)					
Scope of the	.367	.364	.178	1.008	.317
vocabulary					
Vocabulary	-.013	.333	-.006	-.038	.970
command					
(writing					
ability)					

In the following section, the previously reported results will be adopted to test the hypotheses of the first two sub-questions. When we take a closer look at the correlation coefficients of the self-assessment coefficient as a construct and its correlation with both the DIALANG Vocabulary Test and the Cambridge General English Online Test, it can be seen that for the former test, the correlation coefficient is $R = .515$, and that the latter yields a correlation of $R = .650$. The correlation between the self-assessment component and the Cambridge General English Online Test is slightly stronger than the correlation between the self-assessment component and the DIALANG Vocabulary Test. This implicates that there are discrepancies in the correlations between the post-test self-assessment component as one construct and the CEFR levels according to either the DIALANG Vocabulary Test or the Cambridge General English Online Test. Therefore, it is valid to reject H_{01} , and to accept H_{A1} . This is in alignment with the predictions of the current thesis. With regard to the testing of the second hypothesis,

the results show that the correlation coefficient of the Likert scales ($R = .464$) and the DIALANG Vocabulary Test is higher than that of the CEFR descriptors ($R = .456$) and the same test. It can be stated that there are discrepancies in the extent to which these self-assessment items are able to predict the CEFR levels on the DIALANG Vocabulary Test. For the Cambridge General English Online Test, a different result manifests itself. The correlation of the CEFR descriptors ($R = .614$) and the Cambridge General English Online Test is higher than that of the Likert scales and the Cambridge General English Online Test. Namely, this correlation is $R = .566$. Either way, the correlation coefficients between both self-assessment instruments and the language tests are divergent. This means that H_{02} can be rejected, and that we can accept the alternative hypothesis H_{A2} . This is in line with the predictions of the results: there are discrepancies in the correlations between the two parts of the self-assessment component and their correlations with the language testing component.

5.3 Demographic Discrepancies regarding the Correlations between the Self-Assessment Component and the Language Testing Component

In the questionnaire, there are five types of items employed that chart demographic data: age, gender, native language, educational background, and use of English. As age is a scale variable, it is not possible to distribute the data of the multiple regression analysis according to age. For all participants, Dutch is the native language, apart from one Dutch/English bilingual participant. Therefore, it has been decided to perform the multiple regression analysis with the selected variables that distribute gender, educational background and use of English into categories. In the following subsections these analyses will be reported on.

5.3.1 Gender

As previously mentioned, 57 female participants and 16 male participants partook in the study. In the following passage of this section, a multiple regression analysis will be carried out in order to test if there are differences in gender regarding the way the self-assessment component as a unified construct is able to predict the CEFR levels the participants will achieve on the DIALANG Vocabulary Test. A selection variable 'gender' is entered to test for differences in gender regarding the multiple regression analysis. The independent variables, i.e. the predictors, are all the seven items that are employed for the self-assessment component, and the dependent variable consists of the CEFR levels according to the DIALANG Vocabulary Test.

Table 26. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test distributed by Gender

Female				Male			
R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate
.591	.350	.259	1.21771	.763	.582	.165	.94399

It can be deduced from Table 25, that for female participants the correlation coefficient R is .591, which indicates a moderate positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .259, which entails that the independent variables account for the CEFR level the female participants may achieve on the DIALANG Vocabulary Test in 25.9% of the cases. This is fairly higher than for male participants (*Adjusted R²* = .259). For male participants the correlation coefficient R is .763, which implies that there exists a strong correlation between the independent variables and the dependent variable.

The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant for female participants, $F(7, 50) = 3.84$, $p < .005$, $R^2 = .35$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test. It could accordingly be stated that, for females, all self-assessment items, combined as one group, are able to predict the CEFR levels of the DIALANG Vocabulary Test. However, the overall regression model is not statistically significant ($p = .336$) for male participants. According to the DIALANG Vocabulary Test, the independent variables can account for a significant amount of variance in the CEFR levels, which means that the self-assessment items, combined as one category, can predict the DIALANG Vocabulary Test CEFR rates.

Table 27. ANOVA Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test distributed by Gender

Female	Male
--------	------

	Sum of	df	Mean	F	Sig.	Sum of	df	Mean	F	Sig.
	Squares		Square			Squares		Square		
Regression	39.859	7	5.694	3.840	.002	8.696	7	1.242	1.394	.336
Residual	74.141	50	1.483			6.238	7	.891		
Total	114.000	57				14.933	14			

For female participants, the unique variance is marginally statistically significant for the Likert scale ratings of listening comprehension ($p = .073$), and for vocabulary command (writing ability) ($p = .089$). For male participants, the unique variance of each of the seven independent variables, or predictors, is statistically significant for the Likert scale ratings of reading ability ($p = .039$). Accordingly, for male participants, this item uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for.

An equivalent multiple regression analysis was carried out for the Cambridge General English Online Test. The independent variables are equal to the previous multiple regression analysis, in that they consist of all the seven items that are employed for the self-assessment component. At this time, the dependent variable consists of the CEFR levels according to the Cambridge General English Online Test.

Table 28. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test distributed by Gender

Female				Male			
R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate
.622	.387	.301	1.62343	.879	.773	.546	1.47285

It can be deduced from Table 27, that for female participants the correlation coefficient R is .622, which indicates a moderate positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .301, which entails that the independent variables account

for the CEFR level the female participants may achieve on the Cambridge General English Online Test in 30.1% of the cases. This is fairly lower than for male participants (*Adjusted R*² = .546). The independent variable is able to predict the CEFR level the male participants may achieve on the Cambridge General English Online Test in 54.6% of the cases. This is a reasonably high percentage of all cases for male participants. Furthermore, for male participants the correlation coefficient *R* is .879, which implies that there exists a strong correlation between the independent variables and the dependent variable.

The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant for female participants, $F(7, 50) = 4.50$, $p < .005$, $R^2 = .39$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the Cambridge General English Online Test. It could accordingly be stated that, for females, all self-assessment items combined as one group are able to predict the CEFR levels of the Cambridge General English Online Test. However, the overall regression model is not statistically significant ($p = .064$) for male participants. The independent variables are not able to account for a significant amount of variance in the CEFR levels according to the Cambridge General English Online Test for male participants. It could accordingly be stated, that all self-assessment items, combined as one group, are not able to predict the CEFR levels of the Cambridge General English Online Test for male participants.

Table 29. ANOVA Results of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test distributed by Gender

	Female					Male				
	Sum of	df	Mean	F	Sig.	Sum of	df	Mean	F	Sig.
	Squares		Square			Squares		Square		
Regression	83.069	7	11.867	4.503	.001	51.748	7	7.393	3.408	.064
Residual	131.776	50	2.636			15.185	7	2.169		
Total	214.845	57				66.933	14			

For male participants, one of the seven individual self-assessment items (scope of the vocabulary) has unique variance in predicting the CEFR levels according to the Cambridge General English Online Test, that the other items did not account for that is marginally

statistically significant ($p = .089$). For females there are two variables for which the unique variance is statistically significant, namely the Likert scale ratings of writing ability ($p = .079$), and quality of the vocabulary (oral proficiency) ($p = .071$).

5.3.2 Educational Background

Overall, the participants are assigned to one out of eight selected educational levels; MBO (graduated) ($n = 12$), HBO (graduated) ($n = 23$), WO (graduated) ($n = 11$), MBO (enrolled) ($n = 1$), HBO (enrolled) ($n = 10$), WO (enrolled) ($n = 9$), secondary school (graduated) ($n = 5$), and not specified ($n = 2$). Supposing that this section would outline all statistical data in a manner similar to the previous section (see 5.3.1), an overwhelming elaboration of the data would come about. That is, providing the reader with tables of all specifications of the multiple regression analyses with the selection variable selecting cases for each of these eight different educational backgrounds would be complicated and questionable. For the sake of providing the reader with a clear overview of the discrepancies due to educational background in the correlations between the self-assessment component as one construct and the CEFR levels on the DIALANG Vocabulary Test and the Cambridge General English Online Test, this section will address correlations and significance in the predictive trait the independent variables possess in relation to the dependent variable in a rather narrative manner. For one participant that is enrolled in MBO, five participants with a secondary school diploma as highest level of education, and two participants that did not specify their educational background, it is not possible to perform the multiple regression analyses with educational background as the selection variable.

Selecting only cases in the multiple regression analysis for participants for whom the educational background is 'MBO (graduated)', there exists a very strong positive correlation ($r = .902$) between the self-assessment component as a construct (independent variables), and the dependent variable, which consist of the CEFR levels according to the DIALANG Vocabulary Test. The self-assessment component is able to predict the CEFR levels according to the DIALANG Vocabulary Test in 48.8% of the cases (*Adjusted* $R^2 = .488$). However, as the ANOVA results indicate, the independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test ($p = .197$) for MBO (graduated). With regard to unique variance of each of the independent variables of the self-assessment component, there is no independent variable that significantly accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test ($p > .05$). The

correlation between the self-assessment component as a construct and CEFR levels according to the Cambridge General English Online Test – the dependent variable – is high ($r = .770$) for the MBO (graduated) cases. *Adjusted R²* is negative in this case, which could imply that implementing a selection variable which only selects cases in which the educational background is ‘MBO (graduated)’ overfits the model. It could be deduced from the ANOVA results that the regression model is not statistically significant ($p = .611$) for the Cambridge General English Online Test, and that none of the independent variables demonstrates statistically significant unique variance.

For HBO (graduated), a high correlation ($r = .727$) between the independent variables and the dependent variable is present. For participants that graduated HBO, the self-assessment component is able to predict the CEFR levels according to the DIALANG Vocabulary Test in 30.9% of the cases (*Adjusted R²* = .309). According to the ANOVA results, the independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test ($p = .073$). Nonetheless, the independent variable ‘scope of the vocabulary’ carries unique variance in predicting the CEFR level according to the DIALANG Vocabulary Test ($p = .014$). For the Cambridge General English Online Test, selecting only the cases in which the educational background is ‘HBO (graduated)’ yields a high correlation of $r = .772$. *Adjusted R²* is .407, which entails that for the selection of participants that graduated HBO, the self-assessment component is able to predict the CEFR levels on the Cambridge General English Online Test the participants will achieve for 40.7% of the cases. The analysis of variance tells us that the regression model is statistically significant, $F(7, 15) = 3.16$, $p < 0.05$. The unique variance is statistically significant for the Likert scale ratings of writing ability ($p = .008$).

The multiple regression analysis with a selection of cases of participants that graduated WO evinces that there is a very high correlation ($r = .990$) regarding the predictability of CEFR levels according to the DIALANG Vocabulary Test on behalf of the self-assessment component. In this particular case, the self-assessment component is able to predict the CEFR levels according to the DIALANG Vocabulary Test in 93.2% of the cases (*Adjusted R²* = .932). The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant for participants that graduated WO, $F(7, 3) = 20.70$, $p < .05$, $R^2 = .93$. Three of the independent variables – Likert scale ratings of oral proficiency ($p = .022$), Likert scale ratings of writing ability ($p = .009$), and quality of the vocabulary (oral proficiency) ($p =$

.028) – carry unique variance in predicting the dependent variable. The unique variance of the Likert scale ratings of listening comprehension is marginally statistically significant ($p = .070$). The correlation between the self-assessment component as a construct and CEFR levels according to the Cambridge General English Online Test – the dependent variable – is high ($r = .704$) for the selected cases. Similar to the multiple regression analysis, with the value of the selection variable being ‘MBO (graduated)’, *Adjusted R²* is negative for the participants that graduated WO as well, which could imply that implementing a selection variable which only selects cases in which the educational background is ‘WO (graduated)’ overfits the model. The ANOVA results evince that the regression model is not statistically significant ($p = .844$). None of the independent variables carries statistically significant unique variance.

Selecting only cases for participants that are enrolled in HBO, the multiple regression analysis demonstrates a very strong correlation ($r = .987$) between the self-assessment component and the CEFR levels according to the DIALANG Vocabulary Test. The self-assessment component is able to predict the CEFR levels according to the DIALANG Vocabulary Test in 88.4% of the cases (*Adjusted R²* = .884). Yet, the independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test ($p = .087$). The unique variance for both the Likert scale ratings of listening comprehension ($p = .077$), and the Likert scale ratings of reading ability ($p = .077$) is marginally statistically significant. The correlation between the self-assessment component as a construct and CEFR levels according to the Cambridge General English Online Test – the dependent variable – is very strong ($r = .813$) for the HBO (enrolled) participants. Furthermore, *Adjusted R²* is negative, which could imply that implementing a selection variable which only selects cases in which the educational background is ‘HBO (enrolled)’ overfits the model. The ANOVA results show that the regression model is not statistically significant ($p = .764$). None of the independent variables demonstrates statistically significant unique variance.

For participants that are enrolled in WO, the multiple regression analysis demonstrates a very strong correlation ($r = .971$) between the self-assessment component and the CEFR levels according to the DIALANG Vocabulary Test. The self-assessment component is able to predict the CEFR levels according to the DIALANG Vocabulary Test in 53.8% of the cases (*Adjusted R²* = .538). The ANOVA results show that the independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test ($p = .466$), and none of the independent variables carries unique variance ($p = > .05$). For

the Cambridge General English Online Test, selecting only the cases in which the educational background is ‘WO (enrolled)’ yields a very strong correlation of $r = .950$. *Adjusted R²* is .223, which entails that for the selection of participants that are enrolled in WO, the self-assessment component is able to predict the CEFR levels on the Cambridge General English Online Test the participants will achieve for 22.3% of the cases. The analysis of variance tells us that the regression model is not statistically significant ($p = .586$). None of the independent variables demonstrates statistically significant unique variance.

5.3.3 Use of English

In the questionnaire, two items are employed with which the participants are able to address their use of English: (1) use of English in daily life and (2) use of English in the workplace or educational setting. As it is possible to draw up a distribution for the selected cases in which the values are ‘yes’, ‘no’, and ‘sometimes’, a clear overview such as in section 5.3.1 could be provided.

5.3.3.1 Daily Life

It can be deduced from Table 29, that for participants that do use English in their daily life, the correlation coefficient R is .777, which indicates a strong positive correlation between the independent variables and the dependent variable. *Adjusted R²* is .404, which entails that the independent variables are able to predict the CEFR level the participants that do use English in their daily life may achieve on the DIALANG Vocabulary Test in 40.4% of the cases. This is fairly higher than for participants that sometimes use English in their daily life (*Adjusted R²* = .210), and somewhat lower than for participants that do not use English in their daily life (*Adjusted R²* = .480). For participants that do not use English in their daily life the correlation coefficient R is .804, which implies that there exists a strong correlation between the independent variables and the dependent variable for this group of participants. For participants that sometimes use English in their daily life, the correlation between the self-assessment component and the CEFR levels on the DIALANG Vocabulary Test is moderate ($R = .644$).

Table 30. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test distributed by Use of English in Daily Life

Yes	No	Sometimes
-----	----	-----------

R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate
.777	.603	.404	1.19506	.804	.646	.480	.86462	.644	.415	.210	.89575

The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant for participants that do use English in their daily life, $F(7, 14) = 3.04$, $p < .005$, $R^2 = .60$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test. For participants that do not use English in their daily life, the multiple regression analysis shows that the overall regression model is statistically significant as well, $F(7, 15) = 3.91$, $p < .005$, $R^2 = .65$. Nonetheless, the overall regression model is not statistically significant ($p = .102$) for participants that sometimes use English in their daily life. The independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test for participants that sometimes use English in their daily life. It could accordingly be stated, that all self-assessment items, combined as one group, are not able to predict the CEFR levels of DIALANG Vocabulary Test for these participants.

For participants that filled in 'yes' to the question if they use English in their daily life, the unique variance of each of the seven independent variables, or predictors, is statistically significant for the Likert scale ratings of listening comprehension ($p = .001$), the Likert scale ratings of reading ability ($p = .024$), and the Likert scale ratings of writing ability ($p = .035$). Accordingly, for participants that do use English in their daily life, these items uniquely account for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. For participants that do not use English in their daily life, neither of the seven individual self-assessment items uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. This implies that the independent variables correlate with each other to such a degree that none of the independent variables, when considered individually, offers any significant amount of unique variance in predicting the dependent variable. For participants that sometimes use English in their daily life, the unique variance is statistically significant for the Likert scale ratings of reading ability ($p = .011$). The unique variance of the Likert scale ratings of oral

proficiency ($p = .096$) and vocabulary command (writing ability) ($p = .081$) is marginally statistically significant.

For the Cambridge General English Online Test, an equivalent multiple regression analysis with use of English in daily life as the selection variable is performed. For participants that do use English in their daily life, the correlation coefficient R is .742, which indicates a strong positive correlation between the independent variables and the dependent variable. *Adjusted R^2* is .326, which entails that the independent variables are able to predict the CEFR level the participants that do use English in their daily life may achieve on the Cambridge General English Online Test in 32.6% of the cases. This is somewhat higher than for participants that do not use English in their daily life (*Adjusted R^2* = .258). For participants that do not use English in their daily life the correlation coefficient R is .703, which implies that there exists a strong correlation between the independent variables and the dependent variable for this group of participants. For participants that sometimes use English in their daily life, the correlation between the self-assessment component and the CEFR levels on the Cambridge General English Online Test is moderate ($R = .613$), and *Adjusted R^2* is .158, which entails that the independent variables are able to predict the CEFR level the participants that sometimes use English in their daily life may achieve on the Cambridge General English Online Test in 15.8% of the cases.

Table 31. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test distributed by Use of English in Daily Life

Yes				No				Sometimes			
R	R^2	Adjusted	Std.	R	R^2	Adjusted	Std.	R	R^2	Adjusted	Std.
		R^2	Error of the Estimate			R^2	Error of the Estimate			R^2	Error of the Estimate
.742	.551	.326	1.75494	.703	.494	.258	1.53419	.613	.376	.158	1.53959

The ANOVA results of the multiple regression analysis show that the overall regression model is not statistically significant for participants that do use English in their daily life ($p = .073$), for participants that do not use English in their daily life ($p = .109$), and for participants that sometimes use English in their daily life ($p = .160$). The independent variables are not able to

account for a significant amount of variance in the CEFR levels according to the Cambridge General English Online Test for all three of these values of the selection variable. It could accordingly be stated, that all self-assessment items, combined as one group, are not able to predict the CEFR levels of the Cambridge General English Online Test for all three values of use of English.

For participants that filled in 'yes' to the question if they use English in their daily life, the unique variance of each of the seven independent variables, or predictors, is statistically significant for the Likert scale ratings of reading ability ($p = .049$). Accordingly, for participants that do use English in their daily life, this item uniquely accounts for variance in the CEFR levels according to the Cambridge General English Online Test, that the other items did not account for. For participants that do not use English in their daily life, the unique variance of one of the seven individual self-assessment items is marginally statistically significant. This variable is the Likert scale ratings of reading ability, with a p-value of .095. For participants that sometimes use English in their daily life, none of the predictors carries statistically significant unique variance.

5.3.3.2 Workplace or educational setting

In Table 31, it could be noticed that participants that do use English in the workplace or educational setting, the correlation coefficient R is .638, which indicates a moderate positive correlation between the independent variables and the dependent variable. *Adjusted R^2* is .111, which entails that the independent variables are able to predict the CEFR level the participants that do use English in their workplace or educational setting may achieve on the DIALANG Vocabulary Test in 11.1% of the cases. This is reasonably lower than for participants that do not use English in their workplace or educational setting (*Adjusted R^2* = .385). For participants that do not use English in their workplace or educational setting, the correlation coefficient R is .751, which implies that there exists a strong correlation between the independent variables and the dependent variable for this group of participants. For participants that sometimes use English in their workplace or educational setting, the correlation between the self-assessment component and the CEFR levels on the DIALANG Vocabulary Test is strong ($r = .707$), and *Adjusted R^2* = .305.

Table 32. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the DIALANG Vocabulary Test distributed by Use of English in the Workplace or Educational Setting

Yes				No				Sometimes			
R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate	R	R ²	Adjusted R ²	Std. Error of the Estimate
.638	.407	.111	1.44695	.751	.565	.385	.90533	.707	.500	.305	.96307

The ANOVA results of the multiple regression analysis show that the overall regression model is not statistically significant for participants that do use English in their workplace or educational setting ($p = .289$). The independent variables are not able to account for a significant amount of variance in the CEFR levels according to the DIALANG Vocabulary Test. For participants that do not use English in their workplace or educational setting, the multiple regression analysis shows that the overall regression model is statistically significant, $F(7, 17) = 3.15$, $p < .05$, $R^2 = .57$. Likewise, the regression model is statistically significant for participants that sometimes use English in their workplace or educational setting, $F(7, 18) = 2.57$, $p = .05$, $R^2 = .05$.

For participants that filled in 'yes' to the question if they use English in their workplace or educational setting, the unique variance of each of the seven independent variables, or predictors, is marginally statistically significant for the Likert scale ratings of listening comprehension ($p = .061$), and the Likert scale ratings of writing ability ($p = .079$). Accordingly, for participants that do use English in their workplace or education setting, these items uniquely do not account for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. For participants that do not use English in their workplace or educational setting, the Likert scale ratings of writing ability ($p = .045$) uniquely accounts for variance in the CEFR levels according to the DIALANG Vocabulary Test, that the other items did not account for. For participants that sometimes use English in their workplace or educational setting, the unique variance is statistically significant for the Likert scale ratings of reading ability ($p = .009$), and vocabulary command (writing ability) ($p = .012$).

For the Cambridge General English Online Test, an equivalent multiple regression analysis with use of English in the workplace or educational setting as the selection variable is performed. In Table 32, it could be noticed that participants that do use English in the workplace or educational setting, the correlation coefficient R is .766, which indicates a strong positive correlation between the independent variables and the dependent variable. *Adjusted R^2* is .379, which entails that the independent variables are able to predict the CEFR level the participants that do use English in their workplace or educational setting may achieve on the Cambridge General English Online Test in 37.9% of the cases. This is somewhat higher than for participants that do not use English in their workplace or educational setting (*Adjusted R^2* = .303). For participants that do not use English in their workplace or educational setting, the correlation coefficient R is .712, which implies that there exists a strong correlation between the independent variables and the dependent variable for this group of participants. For participants that sometimes use English in their workplace or educational setting, the correlation between the self-assessment component and the CEFR levels on the Cambridge General English Online Test is strong ($r = .707$), and *Adjusted R^2* = .293.

Table 33. Model Summary of the Multiple Regression Analysis of the Self-Assessment Component and the CEFR Levels according to the Cambridge General English Online Test distributed by Use of English in the Workplace or Educational Setting

Yes				No				Sometimes			
R	R^2	Adjusted	Std.	R	R^2	Adjusted	Std.	R	R^2	Adjusted	Std.
		R^2	Error of			R^2	Error of			R^2	Error of
			the				the				the
			Estimate				Estimate				Estimate
.766	.586	.379	1.53239	.712	.506	.303	1.36142	.701	.491	.293	1.62589

The ANOVA results of the multiple regression analysis show that the overall regression model is statistically significant for participants that do use English in their workplace or educational setting, $F(7, 14) = 2.83$, $p < .05$, $R^2 = .59$. The independent variables are able to account for a significant amount of variance in the CEFR levels according to the Cambridge General English Online Test. For participants that do not use English in their workplace or educational setting, the multiple regression analysis is not statistically significant ($p = .059$). Likewise, the

regression model is not statistically significant for participants that sometimes use English in their workplace or educational setting ($p = .057$).

The unique variance of the values 'yes' and 'sometimes' of the selection variable ('use of English in the workplace or educational setting') is not statistically significant for any of the independent variables. Accordingly, for participants that do use, and sometimes use English in their workplace or education setting, these items uniquely do not account for variance in the CEFR levels according to the Cambridge General English Online Test, that the other items did not account for. Therefore, it could be stated that the independent variables correlate with each other to such a degree that none of the independent variables, when considered individually, offers any significant amount of unique variance in predicting the dependent variable. For the participants that do not use English in the workplace or educational setting, the unique variance for the Likert scale ratings of reading ability ($p = .095$) and for quality of the vocabulary (oral proficiency) ($p = .098$) are marginally statistically significant.

Based on the results with regard to gender, educational background and language use, the following section will assess the third hypothesis. Furthermore, the hypothesis regarding the main research question will be assessed in the final section of the current chapter. With regard to gender, it can be seen that the correlation coefficient between the self-assessment component and the DIALANG Vocabulary Test is lower for females ($R = .591$) than the correlation coefficient of male participants ($R = .763$). For the Cambridge General English Online Test, a similar pattern was demonstrated. For female participants, a correlation coefficient of R is .622 regarding the self-assessment component and the Cambridge General English Online was reported, whereas for male participants, this is R is .879. However, this could be explained by the female-male ratio. Far less male subjects participated in this study. Therefore, it is not beneficial to state that for male participants, the self-assessment component as a constructs correlates more, and more positively with both tests. Therefore, these results will not be taken into account to reject the null hypothesis, and to accept the alternative hypothesis. These claims are supported by the results of the multiple regression analysis with the implementation of educational background as a selection variable. It can be seen that for all values of the educational background selection variable, higher correlations will be present. This could be due to a smaller proportion of participants, that are taken into the equation in this regression analysis as well. Therefore, no claims could be made in terms of the influence of educational background on the correlation between the self-assessment component and both segments of

the language testing component. This provides no reason to reject the null hypothesis, and to accept the alternative hypothesis.

The correlation coefficient of the self-assessment construct and the DIALANG Vocabulary Test is R is .777 for participants that do use English in their daily life, R is .804 for participants that do not use English in their daily life, and R is .644 for participants that sometimes use English in their daily life. With regard to the correlation between the self-assessment construct and the Cambridge General English Online Test, R is .742 for people that do use English in their daily life, R is .703 for participants that do not use English in their daily life, and R is .613 for participants that sometimes use English in their daily life. Accordingly, the correlations are somewhat different, with the correlation coefficient of people that sometimes use English in their daily life being lower than for people that do and do not use English in their daily life for both tests. When performing the same multiple regression analysis with the DIALANG Vocabulary Test being the dependent variable and the self-assessment construct the independent variable, with use of English in the workplace or educational setting as the selection variable, it could be seen that the correlation coefficient R is .638 for people that do use English in the workplace or educational setting, R is .751 for people that do not use English in the workplace or educational setting, and R is .707 for people that sometimes use English in the workplace or educational setting. For the Cambridge General English Online Test, the correlation coefficient R is .766 for people that do use English in their workplace or educational setting, R is .712 for people that do not use English in the workplace or educational setting, and R is .707 for the participants that sometimes use English in these particular settings. For the latter test, discrepancies in the correlations are present, but these differences are not as considerable as for the DIALANG Vocabulary Test. This supports the claims that no unambiguous conclusion can be drawn from these results. Hence, it is complicated to claim that the null hypothesis for the third hypothesis could be rejected. Therefore, the alternative hypothesis H_{A3} will not be accepted, and the decision has been made to state that there are no demographic discrepancies, i.e. gender, educational level and use of English, regarding the correlation of the self-assessment component and the language testing component.

Although the demographic features generally seem to have not yielded different results in terms of correlations between the self-assessment construct and the language testing component, the response to the hypothesis of the main research question could be elaborated as follows. All in all, it could be stated that there is a significant correlation between CEFR levels of vocabulary

knowledge, according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test, and the post-test self-assessment component. The results of the hypotheses test of the first two hypotheses regarding the two first sub-questions have been shown to be significant. Therefore, H_{ARQ} will be accepted.

6. Discussion

This discussion will primarily outline some findings that the literature study yielded and attempt to link back these insights to the practical contributions that the experiment in this study has provided.

6.1 Theoretical and Practical Implications

Some interesting insights, and theoretical and practical implications have emerged from the theoretical research that was conducted prior to the experiment. When considering both tests, applied in the language testing component of the questionnaire, it could be concluded that the DIALANG Vocabulary Test contains more open-ended questions than the Cambridge General English Online Test, as the latter test solely contains multiple choice items. As multiple choice formats reduce the chance of negative backwash, it was expected that the test takers score better on the Cambridge General English Online Test in general. As English as a Foreign Language (EFL) speaker often experience difficulties with spelling, the DIALANG Vocabulary Test is at a disadvantage in terms of the reported levels of CEFR. This becomes visible in the sense that no items where all participants filled in the correct answer are present in the DIALANG Vocabulary Test, as for the Cambridge General English Online Test, four items have been filled in correctly by all participants. However, the DIALANG Vocabulary Test yields no responses where one or more distractors are chosen over the correct answer, whereas for the Cambridge General English Online Test, five items (items 12, 14, 21, 22, and 25) report the preference of the distractors over the correct answer option. For one item (item 24) on the DIALANG Vocabulary Test, it has been reported that this item might not be well suited for this particular target population. This manifests itself in the divergent provided answers.

After having reported on the qualitative results that the answers to both language tests have yielded, this section continues with the report of the statistical findings that this study has obtained. Precedently to the calculation of the correlations between the self-assessment

component and the language testing component, the internal consistency of the scales was calculated. The internal consistency of both items that lay out the obtained CEFR levels of either the DIALANG Vocabulary Test, or the Cambridge General English Online Test, is comparatively arguable ($\alpha = .603$). The implication that could be drawn from this fact, is that the CEFR levels of both the DIALANG Vocabulary Test and the Cambridge General English Online Test do not measure the construct of ‘vocabulary knowledge’ in an equivalent manner. This is, per contra, not a requirement for this particular thesis. Nonetheless, it was verified that, when merged together, the seven self-assessment items contain very high internal consistency ($\alpha = .912$). As a matter of fact, Cronbach’s Alpha is higher than the ones that were calculated for the Likert scale ratings and the CEFR descriptors separately. Additionally, the correlations between each of the seven self-assessment items with the DIALANG Vocabulary Test and the Cambridge General English Online Test were calculated. The correlations between the seven individual self-assessment items and the CEFR levels of the DIALANG Vocabulary Test range between $.288 < r < .423$. For the Cambridge General English Online Test, this range is slightly higher ($.440 < r < .585$).

The way in which the prediction was performed in this thesis – by looking for the set of predictors that best predict the outcome measure, i.e. the CEFR levels according to a language testing component, it was intended for the predictors to be correlated. There is indeed a predictive relationship between the self-assessment items, either seen individually or as a construct, and the CEFR levels achieved by the subjects. There are also differences between the correlations of the whole self-assessment component and the two tests, which was questioned in the first sub-question. Furthermore, there are also differences in the correlations between both parts of the self-assessment component and the language testing component, which is the essence of the inquiry of the second sub-question. When we run the multiple regression analysis for the DIALANG Vocabulary Test, it appears that for this test, both the Likert scale ratings and the CEFR descriptors are able to predict the results individually. Hereby, the two self-assessment instruments are entered separately as variables. When the self-assessment items are included as one construct in the independent variables, this again produces a significant result. This means that not only the self-assessment items separately are capable to predict the CEFR levels on the DIALANG Vocabulary Test, but all the self-assessment items as one construct as well. For the Cambridge General English Online Test, these results are statistically significant as well. This implicates that both the Likert scale ratings and the CEFR

descriptors separately, and as a construct are able to predict the results on the Cambridge General English Online Test.

It may be considered complicated to give an unambiguous answer to the third sub-question of this thesis. The ANOVA results of the multiple regression analysis with three of the demographic variables as the selection variable, and the DIALANG Vocabulary Test as the dependent variable, show that the regression model is statistically significant for female participants, but not for male participants. For the Cambridge General English Online Test an equivalent result was reported. This could be due to the fact that there are far more female participants ($n = 57$), than male participants ($n = 16$). Therefore, the model could be underfit for male participants. For the DIALANG Vocabulary Test, the implementation of educational background values as the selection variable evinces that for HBO and WO graduates, the regression model is statistically significant. For all other educational background values, the model is not statistically significant. For the two variables that chart use of English, the model, with the DIALANG Vocabulary Test as the dependent variable, is statistically significant for participants that do, and do not use English in their daily life, but not for participants that sometimes use English in their daily life. For the Cambridge General English Online Test, the model was not statistically significant for all values of use of English. For the selection variable, with the Cambridge General English Online Test being the dependent variable, ‘use of English in the workplace or educational setting’, the model has been proven to not be statistically significant for participants that do use English in the workplace or educational setting. However, the model is statistically significant for people that do not, or sometimes use English in these settings. For the Cambridge General English Online Test, the values that indicate that the participants that do use English in the workplace or educational setting, the model is statistically significant, whereas for people that do not use, or sometimes use English in these settings, the regression model is not statistically significant.

Although it may be difficult to draw an unambiguous conclusion about all acquired results – and hereby, for the purpose of answering the main research question, excluding the results of section 5.3 – in this thesis it is still assumed that there is a significant correlation between the CEFR levels of vocabulary knowledge, according to both the DIALANG Vocabulary Test and the Cambridge General English Online Test, and a post-test self-assessment component. Accordingly, the main hypothesis H_{ARQ} can be accepted.

6.2 Future research

As has become clear in the specification of the self-assessment items in the methodology, the CEFR descriptors in this study only include productive skills, namely writing and speaking. Furthermore, one descriptor consists of the description of the scope of the vocabulary. Since the Likert scales of the language skills, which also serve as self-assessment in this study, do contain the reception skills of reading and listening, it can be stated that the two self-assessment procedures do not completely match. However, this was not the aim of this study. As can be seen in the document publicized by the CEFR⁷, the level descriptors of the receptive skills correspond to those of the productive skills. Those of the productive skills have therefore been used as a template, and formulated in such a way that they relate to the productive skills. However, for future research it may also be interesting to adjust the wording of the descriptors so that the receptive skills are also included in the self-assessment using these level descriptors. Not only the application of the self-assessment items in this study may raise some questions, but also the application of two different types of tests. This study assumes that the Cambridge General English Online Test is a proficiency test as it indicates a global level of language proficiency. One may wonder whether the latter test is actually not a placement test after all. This test also provides CEFR levels where a candidate can achieve a combination of two levels (for example A2/B1). The two tests therefore certainly do not match strictly. However, this is not the aim of this study. Neither test necessarily has to match with the other test. The goal is more to see a global relationship between the language skills, i.e. vocabulary, that subjects think they have, and the results they actually achieve on a language test. This study is therefore rather exploratory, so there should be more testing with a larger group of participants (e.g. students and scholars) to really make statements about self-assessment use in for instance educational contexts, or contexts where language testing is applicable. In order to exclude effects of the differences between the two tests, several further tests could be employed in future research.

7. Conclusion

The theoretical investigations of the current study have demonstrated that both BICS, such as vocabulary (the rationale of this study) and CALP (e.g. reading ability) can manifest themselves in the L2, which could explain the fact that both L1 and L2 CALP are associated with a common underlying language skill. This underlying capacity, according to Krashen (1981), helps

⁷ For further reference, see: <http://www.erk.nl/docent/niveaubeschrijvingen/>.

learners to respond to a new language, to demonstrate some kind of general understanding, and to make sense of the unfamiliar (Krashen, 1981: 159). Another model by Krashen that was described in this thesis is the Monitor model. This model accounts for the way subjects execute self-assessment procedures. As Krashen (1982: 18) claims, subjects use this model in discrete point tests, such as a multiple choice vocabulary test. Fill in the blank tasks cause students to focus on linguistic analysis as well. Therefore, it is likely that for the tasks of the tests in this questionnaire, the subjects have used the Monitor as a form of self-correcting. The self-assessment practices in this study are post-facto: they are performed after the fulfillment of the linguistic task. The Likert scale ratings and the CEFR descriptors could be considered as criterion-referenced, as the subjects do not necessarily compare themselves to other speakers (norm-referenced), and they evaluate themselves against straightforward criteria. However, it could be claimed that some subjects might find it difficult to transfer knowledge to usage, and therefore experience complications in accurately self-estimating vocabulary knowledge. Self-assessment is by some researchers perceived as the most accurate form of assessment, as it is based on assumptions on the learners' behalf, and the learners have access to a broad database of their own achievements and skill deficiencies (Nurov, 2000: 8). The claim that self-assessment is considered as a valid assessment procedure is supported by Sedikides and Strube (1997), that state that subjects are seeking to reduce uncertainty about an aspect of the self (Sedikides & Strube, 1997: 213). However, sometimes the way in which people assess their language skills depends on previous language assessment results. This entails that in some cases, some high-achieving learners tend to underestimate themselves, whereas low-achieving learners tend to overestimate their performance (Leach, 2012; Blanche, 1988). The latter item of information is the entire essence of the origin of this thesis. Self-assessment practices are in some contexts namely used in education. That is why it is important to obtain more information on this subject, because the grades of students are at stake. Unfortunately, at the time of carrying out the experiment, this experiment could not be performed with actual students, due to the closing of all schools in the Dutch context. However, the experiment was conducted with 'the general public', and it has delivered some results that could be interesting for the language testing field, or for educational practices either way. At first sight, one might expect to think that demographic variables such as the educational background a person has, might influence the CEFR level they obtain on the test, and perhaps on the self-assessment component as well. Unfortunately, this could not be confirmed due to the small sample size. The general outcome of this study is that the CEFR levels of vocabulary knowledge obtained on both the DIALANG Vocabulary Test and the Cambridge General English Online Test correlate with the post-test

self-assessment component. This indicates that the participants, in general are able to predict their language skills, in terms of vocabulary.

8. References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. A&C Black.
- Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New directions for Teaching and Learning*, 74, 5-16.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment*. Boston: Heinle & Heinle Publishers.
- Baleghizadeh, S., & Masoun, A. (2014). The effect of self-assessment on EFL learners' goal orientation. *Iranian Journal of Applied Linguistics*, 17(1), 25-48.
- Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986.
- Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety research*, 1(2), 77-98.
- Benmostefa, N. (2008). Types of Language Tests. Chetouane, Algeria. Retrieved from: <https://faclettire.univ-tlemcen.dz/assets/uploads/DOCUMENTS/cours%20en%20ligne/4-TYPE-L-BENM.pdf>.
- Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC journal*, 19(1), 75-93.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: the DLIFLC study. *Language Testing*, 7(2), 202-229.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language learning*, 39(3), 313-338.
- Blue, G. M. (1994). Self-Assessment of Foreign Language Skills: Does It Work?. *CLE Working papers*, 3, 18-35.
- Brantmeier, C., & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, 36(3), 456-477.
- Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40(1), 144-160.

- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Clément, R., Gardner, R. C., & Smythe, P. C. (1980). Social and individual factors in second language acquisition. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 12(4), 293.
- Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of cross-cultural psychology*, 9(2), 131-149.
- Cummins, J. (1980). The construct of language proficiency in bilingual education. *Current issues in bilingual education*, 81-103.
- Cummins, J. (2005). Teaching for cross-language transfer in dual language education: Possibilities and pitfalls. In *TESOL Symposium on dual language education: Teaching and learning two languages in the EFL setting* (pp. 1-18). Estambul: Universidad BogaziciTurquía.
- Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge* (p. 150). Cambridge: Cambridge University Press.
- Dandenault, E. J. (1997). *Self-assessment of communicative ability: investigation of a novel tool for ESL learners* (Doctoral dissertation, Concordia University).
- De Jong, J. H. A. L. (2004, February). Comparing the psycholinguistic and the communicative paradigm of language proficiency. In *International Workshop of Psycholinguistic and Psychometric Aspects of Language Assessment in the Common European Framework of Reference for Languages*, University of Amsterdam, The Netherlands.
- Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, 21(1), 31-46.
- Faez, F., Majhanovich, S., Taylor, S. K., Smith, M., & Crowley, K. (2011). The power of “Can Do” statements: Teachers’ perceptions of CEFR-informed instruction in French as a second language classrooms in Ontario. *Canadian Journal of Applied Linguistics*, 14(2), 1.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of educational research*, 59(4), 395-430.
- Gregg, K. R. (1984). Krashen's monitor and Occam's razor. *Applied linguistics*, 5(2), 79-100.
- Ho, H. F., & Huong, C. (2011). A multiple aspects quantitative indicator for ability of English vocabulary: vocabulary quotient. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1), 3.

- Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. *Second language reading and vocabulary learning*, 153-178.
- Hughes, D. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- 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.
- Janssen-van Dieten, A. M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6(1), 30-46.
- Janulevičienė, V., & Kavaliauskienė, G. (2011). Self-assessment of vocabulary and relevant language skills for evaluation purposes. *Coactivity: Philology, Educology/Santalka: Filologija, Edukologija*, 15(4), 10-15.
- Jones, N., & Saville, N. (2009). European language policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics*, 29, 51-63.
- Klatter, J. & Weltens, B. (2017). Studiesucces van hoogopgeleide vluchtelingstudenten in het Nederlandse hoger onderwijs. Onderzoeksrapport. Radboud Universiteit, Nijmegen.
- Krashen, S.D. (1978). Individual variation in the use of the monitor. *Second language acquisition research: Issues and implications*, 175-183.
- Krashen, S.D. (1979). A response to McLaughlin, "The monitor model: Some methodological considerations." In: *Language Learning*, 29, 151-167.
- Krashen, S.D. (1980). The monitor model for adult second language performance. In K. Croft (Ed.), *Readings on English as a second language* (pp. 213-221). Cambridge, Massachusetts: Winthrop Publishers.
- Krashen, S. D. (1981). Bilingual education and second language acquisition theory. *Schooling and language minority students: A theoretical framework*, 51-79.
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Krausert, S.R. (1991). Determining the Usefulness of Self-Assessment of Foreign Language Skills: Post-Secondary ESL Students' Placement Contribution. Ph.D. Dissertation, University of Southern California.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121.

- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied linguistics*, 29(4), 694-716.
- Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education*, 37(2), 137-147.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL quarterly*, 19(4), 673-687.
- MacIntyre, P. D., Noels, K. A., & Clement, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language learning*, 47(2), 265-287.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211-232.
- Moritz, C. E. (1996). Student Self-Assessment of Language Proficiency: Perceptions of Self and Others.
- Moss, P. A. (1994). Can there be validity without reliability?. *Educational researcher*, 23(2), 5-12.
- Nation, P. (1994). *New Ways in Teaching Vocabulary. New Ways in TESOL Series: Innovative Classroom Techniques*. Alexandria, VA: TESOL, Inc.
- North, B., and Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 217-262.
- Nurov, A. (2000). *Self-assessment of foreign language achievement: the relationship between students' self-assessment, teachers' estimates and achievement test* (Doctoral dissertation, Bilkent University).
- Oller, J. W. (1978). The language factor in the evaluation of bilingual education. In: *Georgetown University Round Table on Languages and Linguistics 1978*. Edited by J. E. Alatis. Washington, D.C.: Georgetown University Press.
- Oppenheimer, L. (2004). Perception of individualism and collectivism in Dutch society: A developmental approach. *International Journal of Behavioral Development*, 28(4), 336-346.
- Oscarson, M. (1980). *Approaches to Self-assessment in Foreign Language Teaching*. Strasbourg. Council of Europe. Oxford, England: Pergamon Press.
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading research quarterly*, 42(2), 282-296.

- Read, J. (2000). *Assessing vocabulary*. Cambridge, England: Cambridge University Press.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7(2), 105-126.
- Rea-Dickens, P., & Germaine, K. (1996). *Evaluation*. Oxford: Oxford University Press.
- Resnick, D., & Resnick, L. (1977). The nature of literacy: An historical exploration. *Harvard educational review*, 47(3), 370-385.
- Saito, Y. (2003). The use of self-assessment in second language assessment. *Studies in Applied Linguistics and TESOL*, 3(1).
- Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of personality and social psychology*, 65(2), 317.
- Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In: *Advances in Experimental Social Psychology*, Vol. 29, Mark P. Zanna, ed. San Diego: Academic Press, 209–269.
- Shuy, R. (1976). Problems in assessing language ability in bilingual education programs. *Bilingual Education, Wayne, NJ: Avery Publishing Group*.
- Starfield, S. (1990). Science and language: a new look at some old issues. *South African Journal of Higher Education*, 4(2), 84-89.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, 103(2), 193.
- Tudor, I. (1996). *Learner-centredness as language education*. Cambridge: Cambridge University Press.
- Verkuyten, M. (2009). Self-esteem and multiculturalism: An examination among ethnic minority and majority groups in the Netherlands. *Journal of Research in Personality*, 43(3), 419-427.
- Wesche, M. B., Morrison, F., Ready, D., & Pawley, C. (1990). French immersion: Postsecondary consequences for individuals and universities. *Canadian Modern Language Review*, 46(3), 430-451.

9. Appendix

9.1 Questionnaire: Engelse woordenschatkennis

Engelse Woordenschatkennis
<p>Deze enquête hoort bij een onderzoek naar de Engelse woordenschat van diverse bevolkingsgroepen. De enquête duurt hoogstens 20 minuten en uw antwoorden zijn volledig anoniem.</p> <p>U kunt de enquête maar één keer invullen. De enquête wordt gesloten op 17 april 2020. Vragen die zijn gemarkeerd met een sterretje (*) zijn vereist.</p> <p>Als u vragen hebt over de enquête, kunt u een e-mail sturen naar: [...].</p> <p>We stellen uw inbreng zeer op prijs.</p> <p>*Vereist</p>
Demografische gegevens
<p>Leeftijd *</p> <p>Jouw antwoord</p>
<p>Geslacht *</p> <ul style="list-style-type: none"><input type="radio"/> zeg ik liever niet<input type="radio"/> mannelijk<input type="radio"/> vrouwelijk<input type="radio"/> anders:
<p>Moedertaal *</p> <ul style="list-style-type: none"><input type="radio"/> Nederlands<input type="radio"/> Engels<input type="radio"/> anders:
<p>(Hoogst genoten) opleiding *</p> <p>Kiezen (opties: (1): ik zit momenteel op het VMBO; (2): ik zit momenteel op de universiteit; (3): ik heb een MBO-diploma; (4): ik zit momenteel op het HBO; (5): zeg ik liever niet; (6): ik heb een HBO-diploma; (7): ik zit momenteel op het VWO; (8): ik zit momenteel op HAVO; (9): ik heb een WO-diploma; (10): ik zit momenteel op de MAVO; (11): ik heb het middelbaar onderwijs afgerond; (12): ik zit momenteel op het MBO).</p>
Deel 1

Kies voor de onderstaande vragen de beste optie om de zin/zinnen te voltooien.
<p>1. Kies het woord dat het best in _____ past: It may be possible to _____ damages against a local authority for not taking care of the roads well enough. *</p> <p>Kiezen (opties: (1) claim, (2) ask, (3) sue, (4) bet)</p>
<p>2. Welk woord kan NIET worden toegevoegd aan het woord 'any' om een nieuw woord te maken? *</p> <p>Kiezen (opties: (1) how, (2) one, (3) way, (4) why, (5) where)</p>
<p>3. Kies het woord dat het best in _____ past: I can't _____ with your offer. Anyone would take pizza instead of a soup mix. *</p> <p>Kiezen (opties: (1) oppose, (2) struggle, (3) examine, (4) compete)</p>
<p>4. Kies het woord dat het best in _____ past: Don't wait any longer, you have to strike while the _____ is hot. *</p> <p>Kiezen (opties: (1) iron, (2) gold, (3) steel, (4) metal)</p>
<p>5. Kies het woord dat het best in _____ past: The armed thief shot the sheriff, and injured the _____ standing next to him. *</p> <p>Kiezen (opties: (1) power, (2) shelter, (3) deputy, (4) fortune)</p>
<p>6. Kies het woord dat het best in _____ past: The _____ and butter of my life? I don't know ... The family? And my work. Making a good living, I suppose. *</p> <p>Kiezen (opties: (1) oil, (2) salt, (3) bread, (4) cheese)</p>
<p>7. Kies het woord dat ongeveer hetzelfde betekent als het woord 'consider'. *</p> <ul style="list-style-type: none"> <input type="radio"/> relate <input type="radio"/> regard <input type="radio"/> expect <input type="radio"/> promise
<p>8. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: The future develop_____ of the European Union depends on what the member states want out of Europe. *</p> <p>Jouw antwoord</p>
<p>9. Welk woord past het best in het balkje in de volgende woordenlijst? Good - better - best _____ - worse - worst *</p> <p>Jouw antwoord</p>
<p>10. Schrijf het woord op dat het best in _____ past. Het woord begint met de letter 'v': He is a pathological killer. His first _____ was 30-year-old Tamara Lind, a former girlfriend. *</p>

Jouw antwoord
11. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: I will try to _____press my feelings more openly, but I'm not sure I can. *
Jouw antwoord
12. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: My latest novel was _____jected by the first three publishers, but with the fourth one I got lucky! *
Jouw antwoord
13. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: My employ_____ gave me a bonus for working overtime. *
Jouw antwoord
14. Schrijf het woord op dat het best in _____ past. Het woord begint met de letter 'c': The _____ of her eyes is brown. * Kiezen (opties: (1) oil, (2) salt, (3) bread, (4) cheese)
15. Beantwoord de vraag met één woord. Het woord begint met 'f': What is usually put around a garden to separate one house from another? *
Jouw antwoord
16. Welk woord betekent ongeveer hetzelfde als het woord 'sick'? Dit woord begint met de letter 'i'. *
Jouw antwoord
17. Kies het woord dat hetzelfde betekent als het woord in HOOFDLETTERS in de volgende zin: There are a number of books and videos on the market, but it's still hard to learn 'tai chi' without personal TEACHING. * <input type="radio"/> selection <input type="radio"/> reading <input type="radio"/> adaption <input type="radio"/> instruction
18. Kies het woord dat hetzelfde betekent als 'motion'. * <input type="radio"/> movement <input type="radio"/> watching <input type="radio"/> reacting <input type="radio"/> converting
19. Kies het woord dat hetzelfde betekent als het woord in HOOFDLETTERS in de volgende zin: It has been my most sincere WISH for some time now. *

<ul style="list-style-type: none"> <input type="radio"/> fault <input type="radio"/> desire <input type="radio"/> request <input type="radio"/> purpose
<p>20. Kies het woord dat het best op de puntjes (...) in beide zinnen past: We built a sandcastle. It ... when the waves came. Our plans ... , when the time allowed for completion was changed. *</p> <p>Kiezen (opties: (1) declined, (2) ruined, (3) collapsed, (4) devastated)</p>
<p>21. Kies het woord dat het best in _____ past: The _____ of this factory is increasing. *</p> <p>Kiezen (opties: (1) title, (2) product, (3) output, (4) aim)</p>
<p>22. Welk woord betekent het tegenovergestelde van het woord dat in de volgende zin in HOOFDLETTERS is geschreven? Schrijf dat woord, dat in _____ hoort, hieronder op. Het woord begint met een 'o': On a sunny day I usually go _____ to get some fresh air. Who wants to stay INSIDE anyway? *</p> <p>Jouw antwoord</p>
<p>23. Welk woord betekent het tegenovergestelde van het woord 'known'? Gebruik dat woord voor _____ in de volgende zin. Schrijf het woord hieronder op: That beach was _____ to us, because we lived so far away from it. *</p> <p>Jouw antwoord</p>
<p>24. Schrijf het woord dat het best in _____ past hieronder op. Het woord begint met 'unc...': The seats were rather _____ , and it was not easy to remain sitting down all that time. *</p> <p>Jouw antwoord</p>
<p>25. Kies het woord dat het best op de puntjes (...) in beide zinnen past: It's raining cats and ...s. Treat somebody like a *</p> <p>Jouw antwoord</p>
<p>26. Kies het woord / de woorden dat/die hetzelfde betekent/betekenen als het woord in HOOFDLETTERS in de volgende zin: The hotel has a small pleasant LOUNGE and bar, two terraces (one on the roof), a solarium (payable locally) and a sauna (free). *</p> <ul style="list-style-type: none"> <input type="radio"/> restaurant <input type="radio"/> discotheque <input type="radio"/> hallway <input type="radio"/> public room

27. Schrijf het ontbrekende deel van het woord met het balkje hieronder op: The Guggenheim Museum is being hailed as the greatest architectural master_____ of this century. *

Jouw antwoord

28. Schrijf het woord op dat het best in _____ past. Het woord begint met de letter 'a': He was badly injured and they took him to hospital in an _____. *

Jouw antwoord

29. Kies het woord dat het tegenovergestelde van 'talkative' betekent. *

- ☐ audible
- ☐ loud
- ☐ quiet
- ☐ mild

30. Welk woord betekent het tegenovergestelde van het woord dat in de volgende zin in HOOFDLETTERS is geschreven? Schrijf dat woord hieronder op: But the state should go further to DISCOURAGE impressionable children from smoking, says political activist Steven Brown. *

Jouw antwoord

Einde deel 1

Hoe goed denkt u dat de vorige opdrachten uw woordenschat in kaart brengen? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Deel 2

Kies voor de onderstaande vragen de beste optie om de zin of het gesprek te voltooien.

1. Can I park here? *

- ☐ Sorry, I did that.
- ☐ It's the same place.
- ☐ Only for half an hour.

2. What colour will you paint the children's bedroom? *

- ☐ I hope it was right.
- ☐ We can't decide.
- ☐ It wasn't very difficult.

3. I can't understand this email. *

- ☐ Would you like some help?

<ul style="list-style-type: none"> <input type="radio"/> Don't you know? <input type="radio"/> I suppose you can.
<p>4. I'd like two tickets for tomorrow night. *</p> <ul style="list-style-type: none"> <input type="radio"/> How much did you pay? <input type="radio"/> Afternoon and evening. <input type="radio"/> I'll just check for you.
<p>5. Shall we go to the gym now? *</p> <ul style="list-style-type: none"> <input type="radio"/> I'm too tired. <input type="radio"/> It's very good. <input type="radio"/> Not at all.
<p>6. His eyes were bad that he couldn't read the number plate of the car in front. *</p> <ul style="list-style-type: none"> <input type="radio"/> such <input type="radio"/> too <input type="radio"/> so <input type="radio"/> very
<p>7. The company needs to decide and for all what its position is on this point. *</p> <ul style="list-style-type: none"> <input type="radio"/> here <input type="radio"/> once <input type="radio"/> first <input type="radio"/> finally
<p>8. Don't put your cup on the of the table – someone will knock it off. *</p> <ul style="list-style-type: none"> <input type="radio"/> outside <input type="radio"/> edge <input type="radio"/> boundary <input type="radio"/> border
<p>9. I'm sorry - I didn't to disturb you. *</p> <ul style="list-style-type: none"> <input type="radio"/> hope <input type="radio"/> think <input type="radio"/> mean <input type="radio"/> suppose
<p>10. The singer ended the concert her most popular song. *</p> <ul style="list-style-type: none"> <input type="radio"/> by <input type="radio"/> with <input type="radio"/> in <input type="radio"/> as
<p>11. Would you mind these plates a wipe before putting them in the cupboard? *</p>

<ul style="list-style-type: none"> <input type="radio"/> making <input type="radio"/> doing <input type="radio"/> getting <input type="radio"/> giving
<p>12. I was looking forward at the new restaurant, but it was closed. *</p> <ul style="list-style-type: none"> <input type="radio"/> to eat <input type="radio"/> to have eaten <input type="radio"/> to eating <input type="radio"/> eating
<p>13. tired Melissa is when she gets home from work, she always makes time to say goodnight to the children. *</p> <ul style="list-style-type: none"> <input type="radio"/> Whatever <input type="radio"/> No matter how <input type="radio"/> However much <input type="radio"/> Although
<p>14. It was only ten days ago she started her new job. *</p> <ul style="list-style-type: none"> <input type="radio"/> then <input type="radio"/> since <input type="radio"/> after <input type="radio"/> that
<p>15. The shop didn't have the shoes I wanted, but they've a pair specially for me. *</p> <ul style="list-style-type: none"> <input type="radio"/> booked <input type="radio"/> ordered <input type="radio"/> commanded <input type="radio"/> asked
<p>16. Have you got time to discuss your work now or are you to leave? *</p> <ul style="list-style-type: none"> <input type="radio"/> thinking <input type="radio"/> round <input type="radio"/> planned <input type="radio"/> about
<p>17. She came to live here a month ago. *</p> <ul style="list-style-type: none"> <input type="radio"/> quite <input type="radio"/> beyond <input type="radio"/> already <input type="radio"/> almost
<p>18. Once the plane is in the air, you can your seat belts if you wish. *</p>

<ul style="list-style-type: none"> <input type="radio"/> undress <input type="radio"/> unfasten <input type="radio"/> unlock <input type="radio"/> untie
<p>19. I left my last job because I had no to travel. *</p> <ul style="list-style-type: none"> <input type="radio"/> place <input type="radio"/> position <input type="radio"/> opportunity <input type="radio"/> possibility
<p>20. It wasn't a bad crash and damage was done to my car. *</p> <ul style="list-style-type: none"> <input type="radio"/> little <input type="radio"/> small <input type="radio"/> light <input type="radio"/> mere
<p>21. I'd rather you to her why we can't go. *</p> <ul style="list-style-type: none"> <input type="radio"/> would explain <input type="radio"/> explained <input type="radio"/> to explain <input type="radio"/> will explain
<p>22. Before making a decision, the leader considered all of the argument. *</p> <ul style="list-style-type: none"> <input type="radio"/> sides <input type="radio"/> features <input type="radio"/> perspectives <input type="radio"/> shades
<p>23. This new printer is recommended as being reliable. *</p> <ul style="list-style-type: none"> <input type="radio"/> greatly <input type="radio"/> highly <input type="radio"/> strongly <input type="radio"/> readily
<p>24. When I realised I had dropped my gloves, I decided to my steps. *</p> <ul style="list-style-type: none"> <input type="radio"/> retrace <input type="radio"/> regress <input type="radio"/> resume <input type="radio"/> return
<p>25. Anne's house is somewhere in the of the railway station. *</p> <ul style="list-style-type: none"> <input type="radio"/> region

- ☐ quarter
- ☐ vicinity
- ☐ district

Einde deel 2

Hoe goed denkt u dat de vorige opdrachten uw woordenschat in kaart brengen? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Kennis van het Engels

Tot slot graag het verzoek om de volgende vragen in te vullen.

Gebruikt u Engels in het dagelijks leven? *

- ☐ ja
- ☐ nee
- ☐ soms

Gebruikt u Engels op uw werk/opleiding? *

- ☐ ja
- ☐ nee
- ☐ soms

In hoeverre bent u het eens met de volgende stelling?: Ik vind het Engels een mooie taal. *

	1	2	3	4	5	
Helemaal oneens	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Helemaal eens

Hoe goed vindt u dat u het Engels begrijpt? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Hoe goed vindt u dat u het Engels spreekt? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Hoe goed vindt u dat u in het Engels kunt lezen? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Hoe goed vindt u dat u in het Engels kunt schrijven? *

	1	2	3	4	5	
Heel slecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Heel goed

Praten in het Engels: kwaliteit van de woordenschat

Hoe zou u uw Engels taalgebruik over het algemeen omschrijven aan de hand van deze situaties? Kies een van de zes situaties. *

Situatie 1. In gesprekken gebruik ik de meest eenvoudige Engelse woorden. Ik kan alleen over alledaagse situaties praten. Mijn zinnen zijn kort en eenvoudig.

Situatie 2. In gesprekken gebruik ik eenvoudige Engelse woorden. Ik kan ook een beetje over minder alledaagse situaties praten. Mijn zinnen zijn kort en eenvoudig.

Situatie 3. In gesprekken gebruik ik woorden zoals de gemiddelde Engelsman die gebruikt. Als ik over mijn vak of interesse praat, kan ik ook redelijk complexe woorden en zinnen gebruiken.

Situatie 4. In gesprekken gebruik ik redelijk complexe Engelse woorden. Spreekwoorden vind ik nog moeilijk. Ik kan redelijk complexe zinnen maken, ook over nieuwe situaties.

Situatie 5. In gesprekken gebruik ik complexe Engelse woorden. Ook gebruik ik spreekwoorden en impliciete taal. Ik kan complexe zinnen maken.

Situatie 6. In gesprekken gebruik ik (zeer) complexe Engelse woorden. Ook gebruik ik spreekwoorden en impliciete taal. Daarnaast begrijp ik dialecten en woorden over onbekende onderwerpen. Ik kan (zeer) complexe zinnen maken.

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6

Bereik van de woordenschat

Hoe zou u uw Engels taalgebruik over het algemeen omschrijven aan de hand van deze situaties? Kies een van de zes situaties. *

Situatie 1. Mijn woordenschat is zeer klein en bestaat uit losse woorden en eenvoudige uitdrukkingen. Met deze woorden kan ik alleen over mijn persoonlijke gegevens en alledaagse situaties praten.

Situatie 2. Mijn woordenschat is klein. Met deze woorden kan ik redelijk goed praten over situaties en onderwerpen die voor mij onbekend zijn.

Situatie 3. Mijn woordenschat is gemiddeld. Met deze woorden kan ik uitgebreid praten over situaties en onderwerpen die voor mij bekend zijn.

Situatie 4. Mijn woordenschat is redelijk groot. Met deze woorden kan ik niet alleen over bekende situaties en onderwerpen praten, maar ook over mijn vakgebied en algemene onderwerpen. Ik kan variatie aanbrengen in mijn woordgebruik, maar vaak aarzel ik als ik een woord niet weet.

Situatie 5. Mijn woordenschat is groot, waardoor ik, als ik een woord niet weet, gemakkelijk andere woorden kan inschakelen. Af en toe moet ik even naar een woord zoeken. Spreekwoorden beheers ik goed.

Situatie 6. Mijn woordenschat is zeer groot en ik kan prima op de woorden komen die ik wil gebruiken. Spreekwoorden beheers ik zeer goed.

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6

Beheersing van de woordenschat

Hoe zou u uw Engels taalgebruik over het algemeen omschrijven aan de hand van deze situaties? Kies een van de vijf situaties. *

Situatie 1. Als ik in het Engels schrijf, kan ik een klein aantal woorden gebruiken die gaan over het dagelijks leven.

Situatie 2. Als ik in het Engels schrijf, kan ik een redelijk groot aantal woorden gebruiken. Als ik schrijf over onbekende situaties en onderwerpen, maak ik soms wel grote fouten.

Situatie 3. Als ik in het Engels schrijf, gebruik ik meestal precies de goede woorden. Als ik niet de juiste woorden gebruik, is mijn geschreven tekst nog wel goed te begrijpen.

Situatie 4. Als ik in het Engels schrijf, heb ik een grote woordenschat. Heel af en toe maak ik een kleine vergissing, maar geen echte fouten.

Situatie 5. Als ik in het Engels schrijf, heb ik een zeer grote woordenschat. Ik maak nauwelijks of geen fouten. Ik kan ook goed betekenisuances aangeven en dubbelzinnigheid vermijden.

<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Einde van de enquête.
Bedankt voor uw deelname!

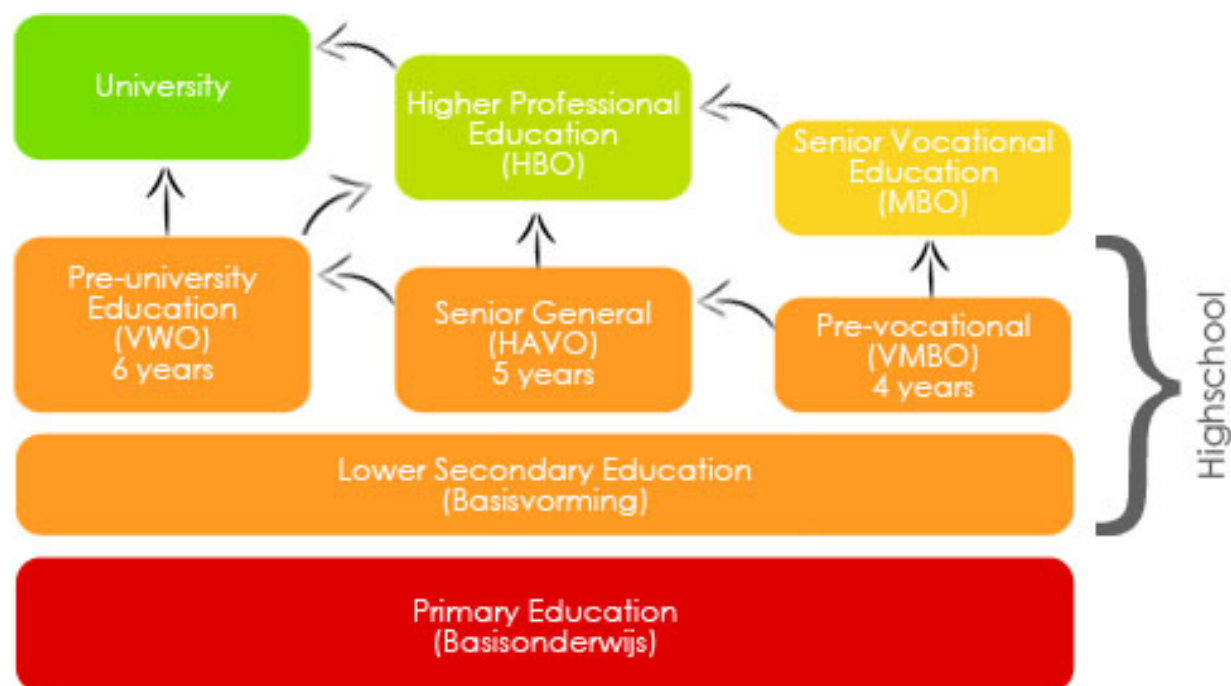
9.2 Answer Grid: Dialang Vocabulary Test and Cambridge General English Online Test

Table 34. Correct answer solutions for each question in both the DIALANG Diagnostic Vocabulary Test and the Cambridge General English Online Test

DIALANG Vocabulary Test		Cambridge General English Online Test	
<i>Item Number</i>	<i>Correct answer option</i>	<i>Item Number</i>	<i>Correct Answer Option</i>
1.	claim	1.	Only for half an hour.
2.	why	2.	We can't decide.
3.	compete	3.	Would you like some help?
4.	iron	4.	I'll just check for you
5.	deputy	5.	I'm too tired.
6.	bread	6.	so
7.	regard	7.	once
8.	ment; -ment	8.	edge
9.	bad; bad-; bad –	9.	mean
10.	victim	10.	with
11.	ex	11.	giving
12.	re	12.	to eating
13.	er; -er; ers; -ers	13.	No matter how
14.	colour; color	14.	that
15.	fence; fences; fencing; a fence	15.	ordered
16.	ill	16.	about

17.	instruction	17.	almost
18.	movement	18.	unfasten
19.	desire	19.	opportunity
20.	collapsed	20.	little
21.	output	21.	explained
22.	out; outside	22.	sides
23.	unknown	23.	highly
24.	uncomfortable; uncomfy	24.	retrace
25.	dog	25.	vicinity
26.	public room		
27.	masterpiece; piece; work; masterwork		
28.	ambulance; automobile		
29.	quiet		
30.	encourage		

9.3 The Dutch Educational System



The Dutch educational system. (n.d.). Retrieved from <https://connect-int.org/connections/2014/08/an-education-on-the-dutch-school-system/>.