Word derivational knowledge and writing proficiency: How do they link?

Although word derivational (WD) knowledge, i.e., how new words are formed from existing words with help of derivational affixes, is considered important for learners of second or foreign languages (L2), there is still no clear answer as to what aspects comprise the construct of L2 English word derivational knowledge and how it develops. The present study adds to our knowledge on how the ability to derive English words develops among L2 English learners. More specifically, it sheds light on how word derivational knowledge relates to communicatively defined Common European Framework of Reference (CEFR) language proficiency levels regarding learners' writing skills. In the study, 117 10th grade learners of English in Estonia and Finland were administered two writing tasks as well as nine measures which were hypothesised to tap learners' word derivational knowledge. The findings indicated that the learners' performance on almost all WD measures were significantly and fairly strongly (at .4–.6 level) correlated with their writing proficiency. The findings also suggest that some aspects of WD ability develop rather steadily between CEFR levels, but others may increase more rapidly after level A2 or B1. These findings thus demonstrate a relationship between word derivational knowledge and language proficiency.

Keywords: *word derivation, L2 proficiency, CEFR, L2 writing*

1. Introduction

Studies that combine language testing and second language acquisition (SLA) research have become more common in the past few decades (e.g. Glaboniat et al. 2005; Bartning, Martin, & Vedder, 2010; Carlsen, 2013; see also Bachman and Cohen, 1998). One reason for this development is the introduction of the Common European Framework of Reference, CEFR, (Council of Europe, 2001). The development of CEFR has created an interest in Europe in how language learners' communicative ability in a foreign or second language (L2), as described in the CEFR levels, develops in terms of linguistic elements of proficiency, that is, vocabulary and structures (Bartning, Martin, & Vedder, 2010). Some of the questions that arose in relation to CEFR included finding out whether the CEFR levels can be distinguished with reference to particular linguistic features or their combinations or to what extent such patterns of linguistic features might depend on learners' first language (L1) or the language they are learning. An interest in finding answers to such questions has characterised the work of several projects across Europe and across several languages such as English (English Profile; e.g., Green, 2012; www.englishprofile.org), German (Profile Deutsch; Glaboniat et al., 2005), and Norwegian (Norsk profil; Carlsen, 2013). The European-wide SLATE (Second Language Acquisition and Language Testing in Europe; www.slate.eu.org) network brings together researchers who share an interest in examining the linguistic basis of the CEFR.

The CEFR has become central to European language education, and it is widely used for setting targets for language learning in curricula and for describing the level of language courses, textbooks and tests (Huhta, 2012; Martyniuk & Noijons, 2007). CEFR levels are also used for such high-stakes purposes as defining language proficiency requirements for citizenship (Extra, Spotti, & van Avermaet, 2009). Despite its widespread

use, the CEFR has been criticised, for instance, for its uncertain basis on second language acquisition research. The framework scales that appear to describe stages of L2 development are not based on empirical research on how proficiency actually develops (Hulstijn, 2007). These criticisms notwithstanding, the fact that the CEFR does not describe the use of any particular language but a language in general means that there is a need to understand how learners coming from a particular L1 background develop in linguistic terms in a particular L2 they are learning.

Word derivation (WD) is a linguistic feature that has received relatively little attention is SLA research so far. Word derivation is the process of forming new words on the basis of existing words, such as *lucky*, *unlucky* and *luckless* from *luck.* It involves the addition of a morpheme such as a prefix or a suffix or both (in the above examples *un-* is an example of a prefix and *-y* and *-less* are examples of suffixes), or an infix (e.g., *Tenne-bloody-see*), which is very rare in English. It should be noted that derivation produces new lexemes and thus differs from inflection which produces grammatical variants of the same lexeme (e.g., *luckier, luckiest*).

The present study adds to our knowledge on how the ability to derive English words develops among L2 English learners. More specifically, we aim at shedding light on how word derivational knowledge relates to CEFR levels defined with reference to learners' writing skills.

Below we will first describe the nature of vocabulary and word derivational knowledge and then present a review of research on derivation and its development, after which we will introduce the current study.

2. Multidimensional and incremental nature of word derivational knowledge

Knowing a word can be defined in several ways. Different lexical models have been presented by, for example, Milton & Fitzpatrick (2013), Nation (2001) and Ringbom (1987). These models can be broadly classified as either dimensional or developmental (see, e.g., Read, 2000, for a discussion). In the following two sections, we will define the two approaches and outline research proposing a) multidimensional and b) incremental models of lexical development.

2.1Multidimensional nature of vocabulary and word derivational knowledge

The first approach to defining vocabulary knowledge seems to be influenced by the connectionist epistemology (e.g., Seidenberg & Gonnerman, 2000), according to which the development of L2 lexical knowledge happens in several knowledge domains, such as orthography, phonology, syntax, and semantics. It dates back to Richards' (1976: 83) influential discussion of the possible dimensions of lexical competence, i.e., knowledge of associations, syntactical properties of words, their form (including derivatives), constraints of use, among others.

One of the well-known dimensional vocabulary knowledge models has been proposed by Nation (e.g., 2001), who outlined three broad aspects of vocabulary knowledge, i.e., form, meaning, and use, and further classified them into subcomponents, e.g., spoken, written, and word parts in the *form* component, as well as differentiated between receptive and productive knowledge of these subcomponents. Ringbom's (1987; 1990) model of lexical knowledge (see **Figure 1**) is similar to Nation's (2001) model. The difference is that it also incorporates the development within each dimension. The developmental approach will be discussed in more detail in the following section.

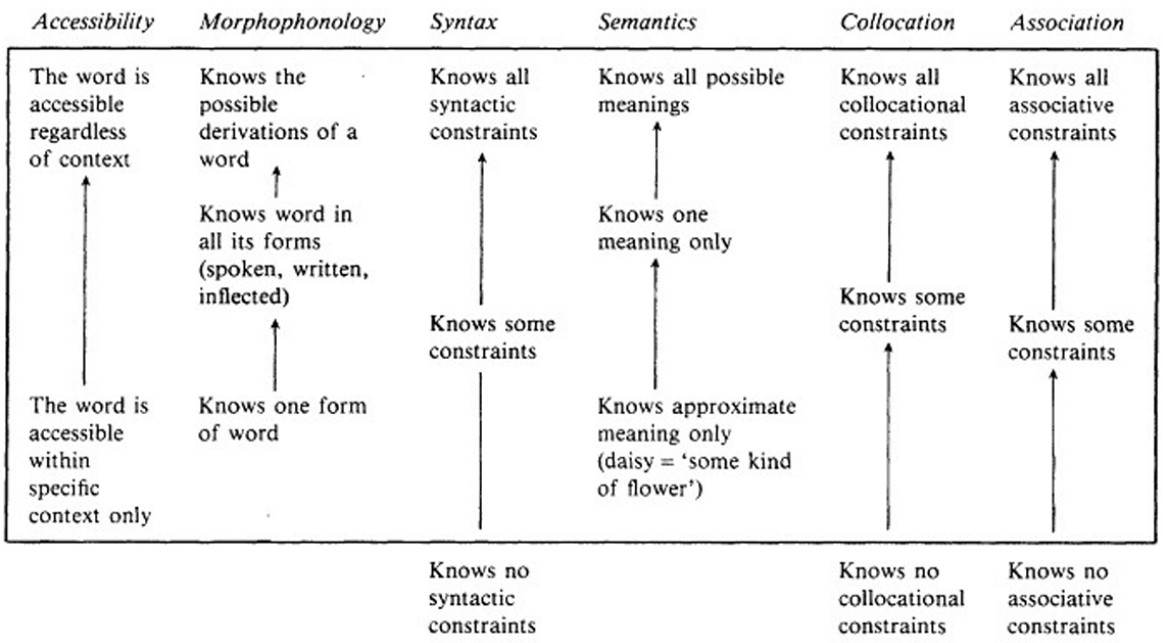


Fig. 1. Ringbom's (1987) model of lexical knowledge.

No comprehensive dimensional model of word derivational knowledge appears to exist. However, research on L2 (and L1) English word derivational knowledge has found that many of the dimensions listed in the vocabulary knowledge models above are also relevant to WD knowledge. These include, for example, syntactic knowledge (e.g., Schmitt, 1998; Schmitt & Meara, 1997; Schmitt & Zimmerman, 2002), knowledge of semantics of derivational affixes (e.g., Chuenjundaeng, 2006), and L1/L2 morphophonology / morpho-orthography (e.g., Alegre & Gordon, 1999; Friedline, 2011). Another dimension is accessibility/control, which has also been labelled as productive/receptive knowledge, or recognition/recall of vocabulary (e.g. Schmitt & Meara, 1997; Hayashi & Murphy, 2010).

2.2 Incremental development of vocabulary and word derivational knowledge

An alternative approach to defining vocabulary knowledge is the developmental one. As the name suggests, this approach stresses development and developmental stages. Research has shown that vocabulary knowledge develops incrementally and correlates

positively with learners' proficiency (e.g., Nation, 2001; Schmitt, 1998; 2010). Similarly, learners' word derivational knowledge appears to develop incrementally, both in L1 and L2 English. For example, Tyler and Nagy (1989) found that while at grade four, learners were able to recognise frequent L1 English stems and derivatives, by grade eight, they increased their syntactic knowledge of derivational affixes. Later, Nagy, Diakidoy, and Anderson (1993) found significant differences in the knowledge of the meaning of frequent L1 English derivational affixes between grade four and upper-secondary school, most of the improvement occurring between grade two and seven.

The development of word derivational knowledge in L2 English acquisition is far less studied than in L1 English acquisition but the findings are similar. For example, Schmitt and Meara (1997) found that university students increased their knowledge of some derivational affixes after one academic year although the increase was modest at best and was not on a par with the increase of their general vocabulary knowledge. In his longitudinal study of four university learners, Schmitt (1998) also found proof for the incremental development of L2 English derivational knowledge although he could not find evidence for any particular order of acquisition.

Not surprisingly, links between learners' knowledge of derivational affixes and their language proficiency and vocabulary size/depth have been discovered although the findings vary. Mäntylä and Huhta (2013) found significant correlations between learners' L2 writing proficiency and their performance on three affix elicitation tasks. Friedline's (2011) cross-sectional study had mixed results as regards L2 proficiency and WD knowledge. Friedline discovered no relationship between language proficiency, as measured by the Michigan Test of English Language Proficiency and learners' performance on a lexical decision task (learners had to state how certain they were that the presented derivative was a real English word) or a word decomposition (learners had to

write the base form of the given derived words) task. However, Friedline's (2011) results suggested that learners' proficiency related to their performance on the word-relatedness task where they had to rate their certainty in the relatedness of the pairs of words, e.g., *decorative–decoration*. Schmitt and Meara (1997) found a moderate correlation ($.27 \leq r \leq .41$) between learners' derivative suffix knowledge, both productive (learners listed all the suffixes that they thought could be attached to the base words given) and receptive (learners marked all the allowable suffixes that they thought could be attached to the base words) and their receptive vocabulary size, as measured by the Vocabulary Levels test (Nation, 1990) but not between their suffix knowledge and language proficiency (i.e., learners' TOEFL scores). The correlation was higher for the receptive suffix knowledge measure. Mochizuki and Aizawa (2010) also found a moderate correlation ($.54 \leq r \leq .65$) between learners' vocabulary size and their knowledge of the meanings of prefixes and syntactic role of suffixes. Hayashi and Murphy (2010) found that the scores on the affix elicitation task were strongly correlated with both productive ($r = .832$) and receptive ($r = .842$) vocabulary size of the Japanese learners of English, but their results on the word segmentation task were not. More recently, Collins and Nation (2015), in an exploratory study, found that learners' ability to understand the meanings of derived words from the previously unknown word families after being provided with the L1 equivalent of the roots did not predict their scores on a vocabulary size test. However, understanding derived forms was a better predictor of their reading speed (in words per minute).

Although following the developmental paradigm, it is tempting to assume that some L2 (and L1) English derivational affixes are acquired earlier than other, their acquisitional order is yet to be discovered—and so is the acquisitional order of WD knowledge dimensions. In this respect, the teaching order of L2 English derivational affixes proposed by Bauer and Nation (1993) and further developed by Nation (2001) could be a starting

point in the process of discovering one. According to them, English derivational affixes could be classified into difficulty levels based on their morphological and phonological properties: frequency, productivity, semantic transparency, regularity of written/spoken form of the bases they attach to, regularity of their spelling/spoken form, and regularity of function. Bauer and Nation (1993) suggested that affixes should be taught to L2 learners in this order. The affix levels as identified by Bauer and Nation (1993) are presented below in **Table 1**.

Table 1.

Teaching order of L2 English derivational affixes (Bauer & Nation, 1993; Nation, 2001).

Level 1	A different form is a different word.
Level 2	Inflectional categories: plural -s, past tense -ed, comparative -er, etc.
Level 3	The most frequent and regular derivational affixes: -able, -er, -ish, -less, -ly, -ness, -th (fourth), -y, non-, un-*
Level 4	Frequent and regular affixes, e.g., -al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-*.
Level 5	Infrequent but regular affixes, e.g., -age -ance, -ship, mis-,etc.
Level 6	Frequent but irregular affixes, e.g., -ee, -ic, -ion, re-, etc.
Level 7	Classical roots and affixes, e.g., -ate, -ure, etc.

*All with restricted uses; see **Appendix 1** in Bauer and Nation (1993) for details.

It should be stressed, though, that Bauer and Nation's (1993) ranking of the affixes by their difficulty is rather arbitrary, and it is premature to consider this an/the order of acquisition. Moreover, to our knowledge, this difficulty order is yet to be corroborated empirically.

Overall, the studies examining learners' L2 English word derivational knowledge are few and result in mixed findings. These studies often consider only a limited number of

dimensions of learners' WD knowledge, which adds to the difficulty of operationalising and generalising the complex construct of learners' L2 English WD knowledge. The present study endeavours to add to the existing body of research by studying whether learners' proficiency relates to their performance on a number of measures estimating hypothesised dimensions of L2 English WD knowledge as represented in Ringbom's (1987; 1990) model of lexical knowledge, and containing derivational affixes from different Bauer and Nation' (1993) levels. We used these levels to introduce variability in the difficulty of affixes, in the absence of an empirically validated order of difficulty.

3. Methodology

3.1 Research questions

In the context of Finnish and Estonian learners of English as a foreign language:

1. Do different aspects of word derivational knowledge relate to learners' writing proficiency?
2. If word derivation and writing are related, is the relationship stable (i.e., do derivation skills increase steadily from level to level) or does ability to derive words increase rapidly at a particular level?

Despite the conflicting results that the previous research on the relationship between learners' L2 English proficiency and word derivational knowledge has produced, informed by Ringbom's (1987; 1990) lexical knowledge model, we hypothesised that different aspects of learners' L2 English word derivational knowledge develop as their proficiency grows.

3.2 Tasks

All in all, two writing tasks as well as nine measures which were hypothesised to tap learners' word derivational knowledge were administered.

Two different writing samples were collected from each learner. The L1 Finnish participants completed the same writing assignments administered as a part of a previous research project in Finland. The writing samples collected from L1 Estonian and Russian participants were a part of their usual classroom assignments, and were thus different for learners at different schools or taught by different teachers. Despite that, the genres / task types were similar in most of the groups, those being argumentative texts (e.g., essays) and formal letters. Other task types, such as narratives (description of an event or a story) were also used. Regarding the genres, it should be noted that judging by the state curricula in Finland and Estonia, learners are expected to write, particularly in the upper-secondary school, and should be familiar with the genres they wrote in. What is more, in the Matriculation Examination in Finland and the English State Exam in Estonia learners are commonly asked to write these types of texts, and these are also covered in different coursebooks used in the schools in the two countries.

The writing samples were independently rated by two raters on the CEFR scale using the procedures and benchmark samples designed in the Finnish research project mentioned above. The ratings were analysed with the multi-faceted Rasch analysis program *Facets* (which we will discuss in more detail when presenting the results), which, to an extent, accounted for the different genres of the written performance samples in different groups. In addition, rating the learners' writing performance on the CEFR scale, which is task-independent, and rating two written performance samples per learner also minimised the possibility of genre affecting the ratings.

The word derivation measures in the study were designed to represent different dimensions of WD knowledge as appearing in Ringbom's model of lexical knowledge (see **Figure 1**). Most of the measures were designed specifically for the current study, as few appropriate measures existed. The lack of measures was particularly acute for measures of word derivation in context; to our knowledge, only Schmitt and Zimmerman (2002) and Mäntylä and Huhta (2013) have developed such measures. We should note that the measures in our study were, nevertheless, somewhat similar to those used in the previous research. Hence, a word segmentation task was also used by Hayashi and Murphy (2010). However, the measure used in the present study involved finding derived words in context rather than using single words as Hayashi and Murphy did. Overall, our measures evaluated both receptive and productive knowledge (active and passive recognition and recall), contained affixes at Bauer and Nation's (1993) levels 3 to 6, formed different parts of speech, and, as regards the base words used as items, belonged to the first five thousand most frequent lemmatised words in the British National Corpus (ref.). The word segmentation task was somewhat different in this respect, as it was based on three excerpts from authentic texts, which we slightly adapted for the purpose of the study. The frequency of the lemmatised items in the task ranged from the first to the twenty-first thousand (the latter being item *revengeful*) most frequent words, with most of the items (k = 31) falling to the first five thousand most frequent words. The items in the word segmentation task were formed with a total of 49 derivational affixes at Bauer and Nation's levels 1-7 (with only 2 items formed with level 7 affixes).

The aim of the word segmentation task (in which the learners were asked to find derived words in three coherent text excerpts and mark derivational affixes in them) was, above all, to study the accessibility dimension of the word derivational knowledge. However, it can be assumed that other types of knowledge, such as semantic and syntactic

knowledge of derivational affixes, were also used by the learners when they worked on the word segmentation task. As to the other measures, the affix elicitation task (in which the learners were asked to form derived words from the words in bold, but also using L1 translations to complete the sentences) regarded the accessibility and semantics dimension. The non-word affix elicitation task (where the learners were required to add affixes to non-words to complete the sentences based on the definitions provided to them) aimed to tap into semantics of derivational affixes and, to an extent, to control for the learners' vocabulary knowledge. In the prefix elicitation task, the learners were asked to select prefixes among provided to complete the derived words in the sentences. The task, we hypothesised, above all, had the semantics dimension (but also, e.g., accessibility). In the grammar recognition task, the learners were required to select one word among the three provided (all having the same bases but different suffixes, forming different parts of speech). The task was to tap into the learners' syntactic knowledge, and so was the aim of the metalinguistic prompts task (although the latter lacked the accessibility dimension and required to demonstrate metalinguistic knowledge). In the metalinguistic prompts task, the learners were asked to write one noun, one verb, and one adjective formed from the given words. The meaning recognition and the passive recognition of the meaning tasks were expected to include the receptive semantics dimension. Finally, the free production task, we suggested, above all, tapped especially into the morpho-phonology dimension. Needless to say, the morpho-phonology dimension was present in all the other measures as well. More details on the measures are presented in **Appendix A**.

All the tasks except for the word segmentation task were administered in an online assessment system, which allowed the participants to complete them faster but also facilitated the coding and the analysis of the data. The afore-mentioned system was designed following the procedure discussed by Fulcher (2003). Specifically, it was

designed to be based on a detailed system and test specifications, and included a multi-stage trialling of the interface and the tasks. For details regarding the description and the usability of the system, see Leontjev (2014).

The WD measures and the instructions, except for the three context-dependent measures (i.e., affix elicitation, non-word affix elicitation, and prefix elicitation tasks), designed and used earlier by Mäntylä and Huhta (2013), were piloted with 22 university students in Estonia (roughly at level B2 of their L2 English proficiency as estimated by their teacher), which allowed us to address some problems in the items (e.g., too difficult items) and in the task instructions. None of the pilot study participants took part in the main study.

3.3 Data and participants

The data come mainly from the learners' performance on the tasks. Additionally, some background data, such as the participants' age, were collected.

The participants in the study were a total of 117 L1 Finnish, Estonian, and Russian learners of English at grade 10, i.e., senior secondary school level (mean age = 16.7; range 15–18), 56 male, 58 female (the sex/gender data of three learners was not available), in Finland and Estonia. There were four different groups of learners taught by two teachers in Finland and five different groups of learners taught by four teachers in Estonia. Further details on the participants are presented in **Table 2**.

Table 2.

Learners' writing proficiency on the CEFR scale.

Country	**N**	**CEFR level***				**Median CEFR level by country**
		A1	**A2**	**B1**	**B2**	

Estonia	47	-	5	22	20	B1
Finland	70	1	11	38	20	B1

*The learners' writing proficiency level is a rounded fair average from Facets (see section 3.2).

The decision to select learners at grade 10 as the participants was rooted in both theoretical and practical considerations. As the study is a part of a larger project, we wanted at least some of the learner participants in the project (who were at grade 10 at the time of the data collection) to participate in the present study. Moreover, according to both the Finnish (Finnish National Board of Education, 2003) and the Estonian (Põhikooli riiklik õppekava õigusakt: Lisa 1, 2010) state curricula, the learners' proficiency in the first foreign language should be at level B1 of the CEFR by the beginning of the senior secondary school (i.e., at grade 10 in both countries), which made the groups more comparable (also see **Table 2**). According to the national curricula, the number of academic hours of L2 English instruction in the first nine years of school in the two countries are somewhat different, that is, 735 in Estonia and 608 in Finland. However, since the data collection in Finland took place about four months later than in Estonia, the amount of instruction in L2 English that the Finnish participants had received was quite comparable to that in Estonia. Furthermore, Nation (2001) suggested that learners can be taught derivational affixes at lower-intermediate level of L2 proficiency, which roughly corresponds to CEFR level B1 (Council of Europe, 2001). It is worth mentioning that we learned from the teachers of the participating learners that they taught their learners word derivation although not extensively and not systematically.

3.4 Procedure

Before the study, the participants granted their permission to use the data for research purposes. As an incentive, they were provided with detailed feedback regarding their performance on the tasks as well as pieces of advice on how to improve their knowledge of vocabulary and word derivation.

A total of two hours was allocated to completing the online tasks measuring the word derivational knowledge, but all learners managed to complete the tasks quicker than that, so the tasks were not speeded. Two groups completed the word segmentation task together with the other tasks (n = 37, in Finland). The rest completed the word segmentation task within a week after the online tasks. When working on the tasks, the learners were in a classroom. In Estonia, a researcher monitored the procedure alongside with the teachers in most of the groups. In Finland, only the teachers did. Detailed instructions were written for the teachers of the participating learners as regards prevention of / reporting on the cases of cheating and responding to the learners' queries during the data collection. The written performance samples were collected within a month and a half before or after the participants completed the WD tasks and these were not speeded either. The written performance samples were checked for plagiarism (also against the work of the other students) to make sure that the learners worked independently on the task.

4 Results

4.1 Reliability of the tasks

Generally, the word derivational knowledge measures used in the study were found to be reliable (internally consistent), for an exploratory study, $.85 \geq \alpha \geq .63$. However, the ten items in the meaning recognition task had a low internal consistency, $\alpha = .46$.

To accompany the reliability analysis, a modern item analysis of the measures was conducted using *Winsteps* Rasch analysis software. The results indicated that item two in

the meaning recognition task was misfitting, *infit MnSq* = 1.57 (*Zstd* = 4.5), *outfit MnSq* = 1.92 (*Zstd* = 4.3). Thus, the task was analysed without this item. Even with this adjustment of the scale, the meaning recognition task had Rasch reliability of .45 and Cronbach's alpha of .56. The grammar recognition and the passive recognition of the meaning task also had somewhat poor Rasch reliability (.63 and .55 respectively), but their Cronbach's alpha coefficients were acceptable for an exploratory study (.73 and .63 respectively). We assume that the main reason for this was the low number of items in the tasks. Therefore, inferences based on the meaning recognition task in particular but also the grammar recognition and the passive recognition of the meaning tasks should be made with caution. The Rasch reliability (ranging from .67 to .93) and the Cronbach alpha coefficients (ranging from .79 to .95) of the rest of the tasks were acceptable. As regards the rater consistency in estimating the learners' proficiency, the Rasch analysis indicated that the ratings were consistent, *infit mean-square* figures being 1 and 0.93 for rater A and B respectively. It should be noted that while the average length of the produced written texts was 130 words, there was a great variation in length (min. = 26; max. = 569). However, even when we controlled for the length, the results of the analyses (see Section 4.2) were interpreted the same.

To reinforce our decision to consider the performance of the two countries (and different L1s) together, we also conducted a Mann-Whitney *U* test on the learners' estimated CEFR proficiency level variable (**Table 2**), the country being the independent variable. The analysis confirmed that the two countries did not differ significantly ($Z = -1.652$, $p = .099$). Moreover, apart from the word segmentation task and the non-word affix elicitation task, the learners in the two countries did not perform significantly differently. The difference in the two tasks was tiny, the country accounting for only 3% to 5% of the variance.

4.2. Word derivational knowledge and writing

In order to address the first research question concerning the relationship between L2 writing and word derivational knowledge, we computed the Spearman rank order correlations between the raw scores from the WD measures and the Facets fair averages based on the ratings of students' writing. Before that, the descriptive statistics are presented (**Table 3**).

Table 3.

Descriptive statistics for the measures used in the study.

Measure	**N**	**Mean**	**S.D.**	**Max. score**	**Total number of items**
Free production task	116	7.65	4.60	29	-
Metalinguistic prompts task	114	5.38	4.62	18	-
Affix elicitation	114	9.58	3.96	15	15
Non-word affix elicitation	114	4.18	3.44	11	13
Prefix elicitation	114	6.64	2.92	12	12
Grammar recognition	113	6.48	2.35	10	10
Meaning recognition	113	5.85	1.93	9	9
Passive recognition of the meaning	111	6.27	2.22	10	10
Word segmentation (# of words)	107	19.39	7.33	36	39
Word segmentation (# of affixes)	106	13.53	5.52	29	49
N listwise	98				

The number of cases differs for different measures because some learners skipped some tasks. The cut-off criterion for considering that the task was skipped was five seconds or less spent on the task. The total possible numbers of correct responses in the free production and the metalinguistic prompts tasks are not indicated in Table 3 and elsewhere in the manuscript. This is because in the free production task, by design, the

number of words the learners were asked to form per item was not limited and in both tasks, learners were allowed to use inflectional affixes as well. **Table 4** presents the correlations for the entire group (number of learners varied from 106 to 117).

Table 4.

Correlation of the word derivation measures with the learners' writing proficiency (Facets fair averages).

Measure	Spearman rho	Significance
Free production task	.458	<.001
Metalinguistic prompts task	.465	<.001
Affix Elicitation	.585	<.001
Non-word affix elicitation	.410	<.001
Prefix elicitation	.581	<.001
Grammar recognition	.642	<.001
Meaning recognition	.578	<.001
Passive recognition of the meaning	.504	<.001
Word segmentation (# of words)	-.101	.300
Word segmentation (# of affixes)	-.001	.989

Correlational analysis of the relationship between writing in English and word derivation measures revealed that some of the latter had strong (over .5 or .6) correlation with writing and even the lower correlations were over .4. The only exception was the word segmentation task in which the learners had to mark in a text all derived words and all affixes that they could find. The number of words or affixes the learners could locate did not correlate at all with their writing proficiency.

To further investigate the relationship between writing proficiency and word derivation, we conducted a multiple linear regression with the Facets fair average for the learners' writing proficiency as the dependent variable and the WD measures as the supposed predictors[2]. The word segmentation task was excluded from the analysis. We

also bootstrapped confidence intervals using Bias-Corrected and accelerated method and 2,000 resamples.

The results indicated that the linear combination of the prefix elicitation task (β = .366, $t(99) = 4.19$, $p < .001$), the grammar recognition task (β = .290, $t(99) = 3.20$, $p = .002$) and the meaning recognition task (β = .246, $t(99) = 3.12$, $p = .002$) significantly related to the learners' writing proficiency ($R^2 = .58$, $R^2_{adj} = .57$, $F(3,99) = 45.49$, $p < .001$), accounting for about 57–58% of the variance[3]. Since the meaning recognition task had a low reliability, we also ran a regression without this measure. This time, the affix elicitation task emerged a significant predictor, too (β = .260, $t(99) = 2.95$, $p = .012$). The variance that these three measures accounted for was similar to that in the first linear regression analysis, $R^2 = .58$, $R^2_{adj} = .56$.

Next, we examined whether learners' derivation ability increases as their writing proficiency grows and whether this increase is steady. For this, we divided the learners into three groups according to their writing proficiency by rounding the students' fair average scores from the Facets analysis into the nearest CEFR level. All learners except one could be placed at A2, B1, or B2 levels. The one student placed at A1 level was included in the nearest, A2, group. We then computed the percent correct scores separately for the three groups. The results presented in **Table 5** reveal that the changes were not that steady across the levels.

Table 5.

Mean percent correct at different CEFR proficiency levels across the measures.

Measure	Proficiency on the CEFR scale		
	A2	**B1**	**B2**
Free production	-	-	-
Metalinguistic prompts	-	-	-
Affix Elicitation	37	61	79

Non-word affix elicitation	14	30	43
Prefix elicitation	24	55	69
Grammar recognition	42	59	83
Meaning recognition	49	59	80
Passive recognition of the meaning	46	60	73
Word segmentation (# of words)	47	52	47
Word segmentation (# of affixes)	25	28	28

To study the differences statistically, we ran a number of ANOVAs, the raw scores in each of the measures being the dependent variable and the CEFR level, the between-subjects independent variable. We then supplemented the regular ANOVA analyses with the linear contrast analyses to establish whether the differences across the learners' proficiency levels related linearly to their WD knowledge as estimated by our measures. In the cases where the homogeneity of variance assumption was violated, we utilised Welche's F-test instead of the regular F-test. The following **Table 6** gives an overview of the results of the ANOVAs we obtained.

Table 6.

Relationship between the learners' CEFR level and their performance on the measures (analyses of variance).

Measure	**F-test / Welche's F-test**	**Effect size**
Free production	$F(2, 113) = 7.20$, $p = .001$	$\eta^2 = .11$
Metalinguistic prompts	Welch's $F(2, 45.29) = 11.66$, $p < .001$	$\eta^2 = .20$
Affix elicitation task	$F(2, 111) = 20.13$, $p < .001$	$\eta^2 = .27$
Non-word affix elicitation[4]	Welch's $F(2, 49.76) = 10.65$, $p < .001$	$\eta^2 = .13$
Prefix elicitation	Welch's $F(2, 36.38) = 22.04$, $p < .001$	$\eta^2 = .34$
Grammar recognition	$F(2, 110) = 35.60$, $p < .001$	$\eta^2 = .39$
Meaning recognition	$F(2, 110) = 22.37$, $p < .001$	$\eta^2 = .29$
Passive recognition of the meaning	$F(2, 108) = 10.59$, $p < .001$	$\eta^2 = .16$
Word segmentation, # of words	$F(2, 104) = 0.71$, $p = .492$	$\eta^2 = .01$
Word segmentation, # of affixes	$F(2, 103) = 0.31$, $p = .731$	$\eta^2 = .006$

The results of the analyses of variance revealed that the effect of the learners' proficiency varied from moderate to strong in all of the measures with the notable exception of the word segmentation task. Additionally, the trend analyses revealed that except for the word segmentation task, the linear trend accounted for most of the variance associated with the learners' writing proficiency. For example, in the meaning recognition task, where there was the biggest difference between the effect of the learners' proficiency and that of the linear trend associated with the proficiency, the linear trend accounted for 24% of the variance (cf. **Table 6**). The linear trend also transpires in the means plots, to which we added 95% confidence intervals (**Appendix B**).

We then followed the ANOVAs with the pairwise comparisons, in which, to account for unequal sample sizes, we used Hochberg's GT2 when the variances were homogeneous and Games-Howell post-hoc tests when they were not[5]. **Table 7** gives a summary of the pairwise comparisons.

Table 7.

Pairwise comparisons of the learners' performance on the measures across the CEFR levels.

Measure	Mean difference and significance		
	A2–B1	**B1–B2**	**A2–B2**
Free production	1.89, $p = .313$	2.55, $p = .016$	4.44, $p = .002$
Metalinguistic prompts	1.39, $p = .278$	3.86, $p < .001$	5.25, $p < .001$
Affix elicitation ($k = 15$)	3.70, $p < .001$	2.62, $p = .001$	6.32, $p < .001$
Non-word affix elicitation ($k = 13$)	2.09, $p = .015$	1.68, $p = .055$	3.77, $p < .001$
Prefix elicitation ($k = 12$)	3.61, $p < .001$	1.77, $p = .005$	5.37, $p < .001$
Grammar recognition ($k = 10$)	1.72, $p = .003$	2.44, $p < .001$	4.16, $p < .001$
Meaning recognition ($k = 10$)	0.94, $p = .118$	1.85, $p < .001$	2.79, $p < .001$
Passive recognition of the meaning ($k = 10$)	1.39, $p = .053$	1.29, $p = .009$	2.68, $p < .001$
Word segmentation (# of words) ($k = 39$)	1.71, $p = .832$	- 1.70, $p = 621$	0.01, $p = 1.00$

Word segmentation (# of affixes) (k = 49)	1.30, p = .831	0.01, p = 1.00	1.31, p = .848

The pairwise comparisons demonstrated that while the trend analysis indicated that the relationship was fairly linear across the measures, the differences between A2 and B1 and B1 and B2 were not as stable. In the affix elicitation, the prefix elicitation and the grammar recognition tasks, both the difference between A2 and B1 and between B1 and B2 were significant (hereinafter in the paragraph, at $p < .05$). In the metalinguistic prompts, the meaning recognition and the passive recognition of the meaning tasks, only the difference between B1 and B2 was significant. The results obtained on the pairwise comparisons for the passive recognition of the meaning are somewhat counterintuitive when percent correct figures across the CEFR levels are considered (see **Table 5**). Nevertheless, the results obtained on the ANOVA should be considered more reliable, as the latter takes into consideration variances and measurement errors. In the non-word affix elicitation task only the difference between level A2 and B1 was significant. It should also be noted that in the free production task, based on the results of the Hochberg's pairwise comparisons, there was a significant difference between levels B1 and B2. However, judging by the means plot with the confidence intervals added (**Appendix B**), only the difference between A2 and B2 was significant.

5. Discussion

The present study aimed at exploring whether learners' L2 English word derivational knowledge is related to their writing proficiency and whether this relation is steady across learners' proficiency levels on the CEFR scale. While comparing learners' proficiency with word derivational knowledge is not new, this study adds to previous research in that we analysed a more comprehensive number of aspects of word derivation.

While the results of the previous research regarding the relation of word derivational knowledge and learners' more general proficiency were mixed (Mäntylä & Huhta, 2013; Friedline, 2011; Schmitt & Meara, 1997), the present study indicated a rather strong correlation between the participants' proficiency and their word derivational knowledge.

The only exception was the word segmentation task, which did not correlate with the learners' writing proficiency. This finding was in line with the previous research, such as Friedline's (2011) and Hayashi and Murphy's studies (2010). We hypothesise that in the present study, this can be attributed to the effect of the task. The learners were asked to find derived words in a coherent text. Thus, they could not but process the text for its meaning. More able learners then were able to recognise more words, which may have interfered with their ability to analyse the words (Ullman, 2001). On the other hand, in Friedline (2011) and Hayashi and Murphy (2010), single words were used in segmentation/decomposition tasks, and not a coherent text. Thus, there seem to be other factors that affect learners' ability to analyse words. Psycholinguistic theories, such as declarative/procedural model (e.g., Ullman, 2001), predict that L2 learners are more dependent on declarative memory and thus are likely to store more L2 linguistic forms as entities in their memory than L1 linguistic forms, can explain this finding. Nevertheless, it is clear that more research into this is required.

The similarity of correlations between WD and writing proficiency in the present study and that of Mäntylä and Huhta (2013) is not surprising, at least for the affix elicitation, the prefix elicitation, and the non-word tasks, as these measures were used in both studies. The strong correlation found by Hayashi and Murphy (2010) between the participants' performance on the affix elicitation task and their vocabulary knowledge is also strengthened by our findings, as vocabulary knowledge is found to correlate with proficiency quite strongly (see, e.g., Schmitt, 2010).

However, the present study also found strong correlations between the learners' syntactic (grammar recognition and metalinguistic prompts tasks, i.e., both receptive and productive) and semantic (meaning recognition and passive recognition of the meaning) knowledge of derivational affixes and learners' proficiency. Interestingly, based on their results obtained on two tasks requiring learners to demonstrate syntactic knowledge of derivational affixes, Schmitt and Zimmerman (2002) also suggested a relationship between the derivational knowledge and proficiency, but did not test their assumption statistically which was done in the present study. On the other hand, there is a discrepancy between our findings regarding the correlation between the proficiency and learners' performance on the free production task and those of, for example, Schmitt and Meara (1997) on a similar task. In the latter study, however, the non-significant correlation might have been due to the small sample size (n = 28).

While the results clearly demonstrated the relation of the participants' writing proficiency and their word derivational knowledge, only some of the measures significantly predicted the learners' writing proficiency. However, they accounted for over 50 percent of the variance.

The likely reason for the grammar recognition task being a significant predictor of writing is that learners constantly refer to their syntactic knowledge when writing in L2, and the latter develops as their abilities in their L2 grow, or at least, the accuracy in and complexity of its use increase (see, e.g., Alanen & Kalaja, 2010). We hypothesise that the reason for the prefix elicitation task significantly predicting the learners' writing proficiency is that it, above all, required learners to refer to their semantic knowledge of the prefixes either as such or through analogy with words containing the same prefix. Also, the prefixes were to a large extent transparent. The meaning recognition task, which also predicted writing though the prediction was much smaller, was designed to tap learners'

semantic knowledge of derivational affixes. Yet, prefixes are, arguably, more transparent, which may be why the prefix elicitation task predicted writing more strongly. As regards the affix elicitation task, it required the participants to demonstrate their vocabulary knowledge more than the rest of the tasks. In fact, Hayashi and Murphy's (2010) finding that learners' vocabulary knowledge strongly predicted their performance on an affix elicitation task indirectly speaks in favour of this interpretation. The question is, however, why other tasks, such as the metalinguistic prompts task, did not predict writing proficiency, although the latter required the learners to recall rather than recognise the affixes and their syntactic functions. We assume that in those tasks, there could have been other factors that interacted with the learners' performance. Specifically, in the case of the metalinguistic prompts task, the learners had to refer to their metalinguistic knowledge, which has been found to relate with language proficiency only weakly (e.g., Alderson, Clapham, & Steel, 1997).

We were also curious as to whether there is a possibility that learners' ability to derive words in English increases more rapidly at a particular level of their proficiency. While overall the linear trend in the measures across the CEFR proficiency levels was rather strong, as can be deduced from both the linear trend analyses and the means plots (**Appendix B**), in roughly half of the measures, there was a bigger change in the learners' performance between levels B1 and B2 than between levels A2 and B1. In two of the measures where it was not the case, that is, the measures also used in Mäntylä and Huhta's (2013) study (the latter arriving to a similar finding to ours), the differences were almost the same across the proficiency levels (see **Table 5** and **Appendix B**). The notable exception was the prefix elicitation task, where there was clearly a bigger increase between A2 and B1.

6. Conclusion

Taken together, the findings suggest that depending on the way word derivational knowledge is operationalised, the ability to derive words is either more or less stable or increases more after level B1. In addition, it seems that syntactic and semantic aspects of word derivational knowledge predict learners' proficiency stronger than others. The results, thus, not only demonstrate that there is a link between word derivational knowledge and writing proficiency, but also suggest that not all of its aspects develop steadily as learners' proficiency grows.

Next, we will discuss the limitations of this study. The first limitation concerns the way we operationalised learners' writing proficiency. The figures were based on two written performance samples (e.g., an essay or a formal letter) per learner rated by two raters only. Moreover, as regards the Estonian sample, at least for some groups, the task types were different. While the approach we selected, i.e., using a Rasch estimation of the learners' abilities, improves the quality of the ratings, having a third rater and asking the learners to complete exactly the same writing tasks would have increased the reliability of the scale representing their writing proficiency.

Moreover, this was a cross-sectional study, so it does not provide an ideal basis for interpreting results in terms of the development of the learners' ability to derive words in English. Rather the findings can be considered a starting point in accumulating evidence for the development of different aspects of word derivation that should be confirmed in longitudinal studies. In fact, the nature and quality of operationalising L2 English proficiency is likely to vary across studies (cf. Friedline, 2011; Schmitt & Meara, 1997), which makes a systematic comparison of the findings with the previous research even more difficult.

While we did not analyse the two countries separately due to the verisimilitude of their performance on the tasks, the participants' L1 might have still influenced the results somewhat. On the other hand, the sample sizes in different L1 groups were even more unequal than those across the proficiency levels, which could have complicated the analyses further and increased the possibility of making both Type I and Type II errors. Nevertheless, we think that future studies of L2 word derivational knowledge should take participants' L1 into consideration. An interesting possibility is that not only learners' mother tongue influences their word derivational knowledge (or some aspects of it) but also their second language(s) do, which could be addressed in a future study. Considering the participants L1's, since Finnish and Estonian are related languages, it is difficult to say whether and to what extent the results are generalisable to learners of other L1s.

Caution should also be exercised when generalising the results of the present study to other measures of word derivational knowledge, as the results could be due to the effect of the task type. Moreover, due to the complex nature of L2 English word derivational (and vocabulary) knowledge, we cannot claim that a certain aspect of word derivational knowledge is more difficult than another despite the results on the pairwise comparisons of the ANOVAs. At the very least, this would require controlling for frequency of the bases, the whole lemmas and their semantic transparency, as well as the length of entire items (e.g., words, phrases, sentences) before we could argue that the task/item difficulty was due to a certain aspect of word derivational knowledge or an interplay of several such aspects.

Finally, as also previous studies on WD have found, it is difficult to separate word derivational knowledge from general vocabulary knowledge. We addressed this challenge by having several different methods of study. Still, the differences in the learners' performance across the CEFR levels and their correlations with the learners' writing

proficiency might have also been due to factors other than their word derivational knowledge. More sophisticated statistical analyses, such as Structural Equation Modelling, could shed more light on this issue in future studies.

Since word derivational skills enhance vocabulary learning, studying and understanding them is worthwhile. Still little researched questions include establishing the effect of the second language and the type of language teaching, doing a more systematic division of different aspects of word derivational knowledge into receptive and productive types, tracing the development of word derivational knowledge incrementally, and confirming and rejecting empirically the order of derivational affixes proposed by Bauer and Nation (1993). Besides that, more research is needed to be able to answer the questions of whether the development of L2 English word derivational knowledge is steady and if not, what point in learners' proficiency can be considered as crucial for its development.

6. Endnotes

1. It should be noted that the average length of the written performance samples in Estonia was significantly higher. However, as an ANCOVA demonstrated, length controlled for, there was no significant difference in the learners' proficiency between the two countries, $F(1, 114) = 2.71$, $p = .102$,$\eta_p^2 = .02$. Moreover, the correlational and the regression analyses conducted with the fair average variable (representing the learners' proficiency) residualised on the average length variable (i.e., the learners' proficiency with the variance introduced by the text length excluded) demonstrated that the figures were similar to those obtained on the original proficiency variable. Thus, regardless of the length of the text, the same conclusions were drawn from the analyses.

2. The figures were obtained using the list-wise deletion of cases. However, the pairwise deletion of cases resulted in the same significant predictors although their order was different.

3. The writing proficiency variable was leptokurtic (i.e., peaked), which is not surprising considering that the sample selected for the study being represented by the learners studying in the same school year. To check whether there was any observable influence of the kurtosis of the DV on the results of the regression analysis, we randomly removed half of the cases at the B1 level (n = 16) from the sample, which resulted in the total of 89 cases in the analysis. The resulting variable was normally distributed (*S-W*(89) = .98, p = .139). The analysis conducted on this variable demonstrated that the same variables and in the same order significantly predicted the writing proficiency (R^2 = .62, R^2_{adj}= .61, F(3, 85) = 45.59, p < .001).

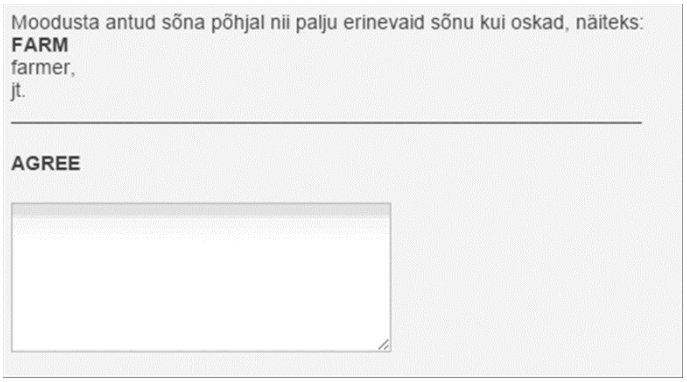
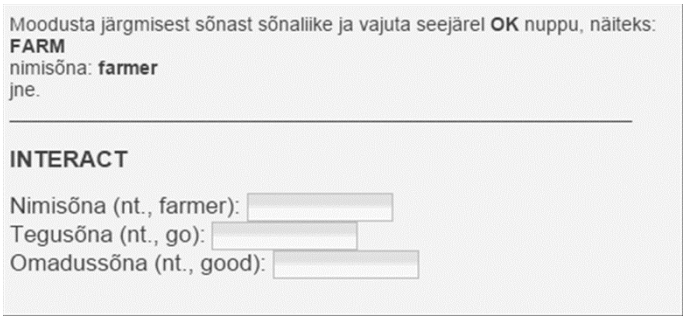
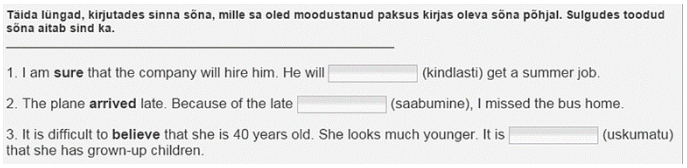
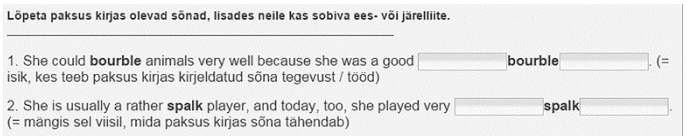
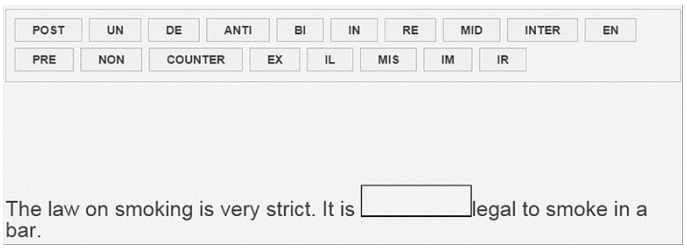
4. About 22% of the learners scored zero on the task, which can be considered a floor effect.

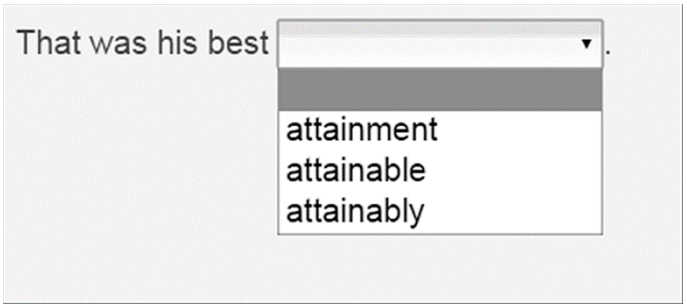
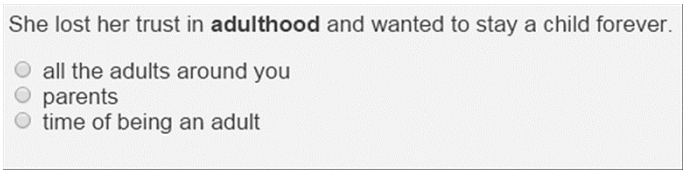
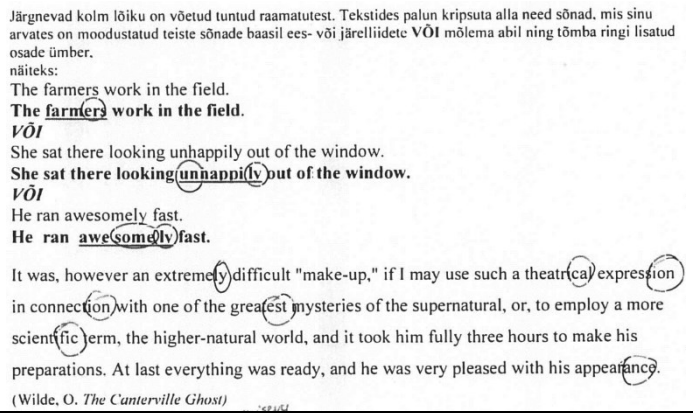
5. In fact, Welch's F-test results might also be not reliable when the distributions are differently skewed, as was the case with the non-word affix elicitation task. However, considering the p-value of the Welch's F-test being less than .001 (in effect, .00014), we think that its result was reliable enough. Thus the post-hoc pairwise comparisons could be conducted.

6. The meaning recognition task had a low reliability, and the affix elicitation task appeared as a significant predictor when the meaning recognition task was excluded.

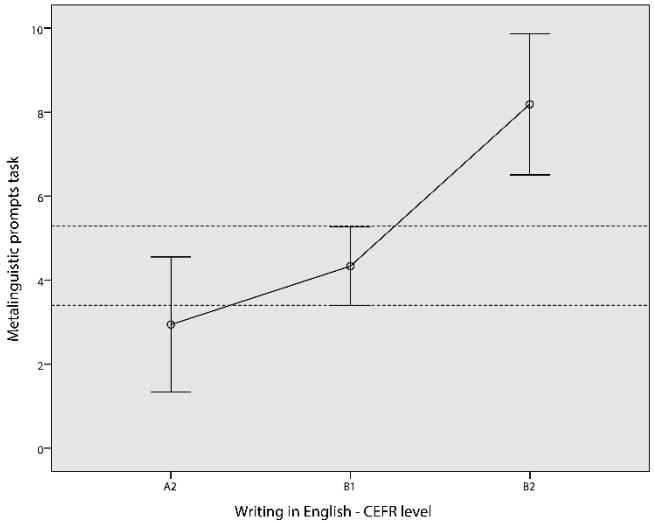
Appendix A. Measures used in / designed for the study and the rationale for their selection.

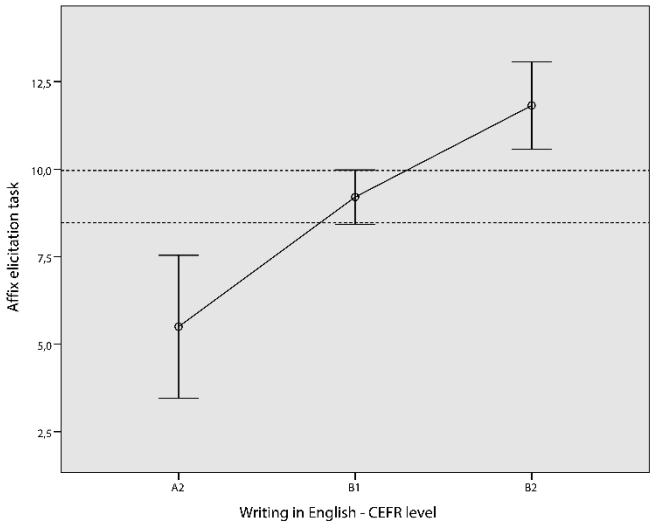
Name of the measure	Description	Sample item

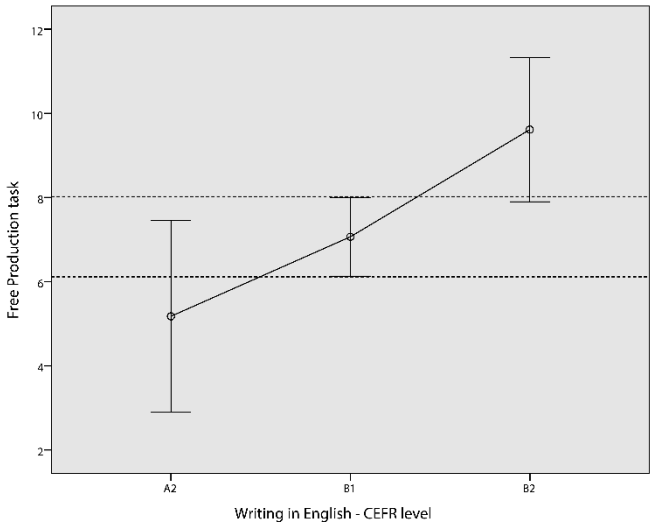
Free production	Ability to produce different derivations formed from the base word (10 items); context-independent (see, Schmitt & Meara, 1997 for a similar measure).	
Meta-linguistic prompts	Ability to produce derived words by using metalinguistic information (names of the parts of speech; 10 items, each requiring producing 3 parts of speech); context-independent; similar to Schmitt (1998)'s measure, although the modality of the latter was oral.	
Affix elicitation	Ability to produce frequent, derived words in context (15 items); context-dependent (see, e.g., Friedline, 2011; Hayashi & Murphy, 2010, for similar measures).	
Non-word affix elicitation	Ability to produce derived forms of non-words (13 items); context-dependent. A modification of the affix elicitation task. The non-words were taken from the list developed for the Vocabulary Size Placement Test of Dialang (Alderson, 2005).	
Prefix elicitation	Ability to produce derived words by supplying the correct prefix, selecting the latter from the provided list (12 items); context-	

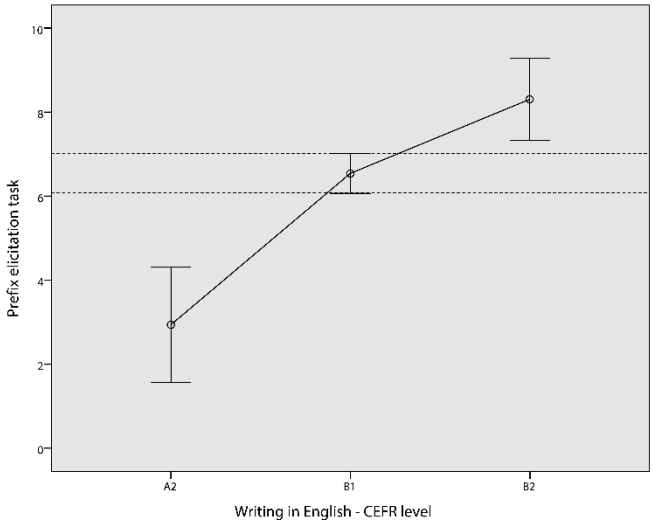
	dependent. A modification of the affix elicitation task.	
Grammar recognition	Ability to recognize correct vs. incorrect derivation in terms of different parts of speech (and meaning); context-dependent (10 items). A measure similar to Akande's (2003) test of knowledge of inflectional affixes.	
Meaning recognition	Ability to recognize correct vs. incorrect derivation in terms of meaning; context-dependent (10 items). A modification of the grammar recognition task adapted for recognition of semantics of derivational affixes, i.e., same bases, different affixes in the options.	
Passive recognition of the meaning	Ability to recognize the meaning of derived words; context-dependent (10 items). Somewhat similar to Nation's (2008) (also Nation & Gu, 2007) Vocabulary Size Test adapted for word derivation, i.e., the options elicited different meanings of derivational affixes.	
Word segmentation	Ability to recognize derived forms / derivational affixes in text context; context-dependent (49 derivational affixes). Similar to Hayashi and Murphy's (2010) word segmentation task (although in the latter, single words were used)	

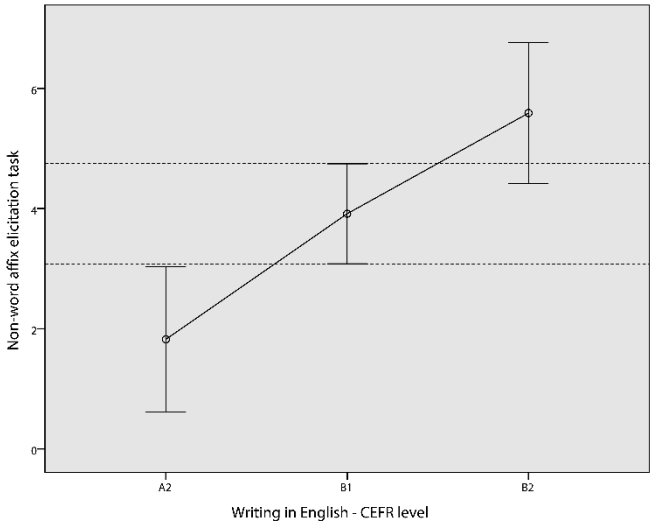
Appendix B. Differences across the CEFR levels in the measures: Means plots with 95% confidence intervals

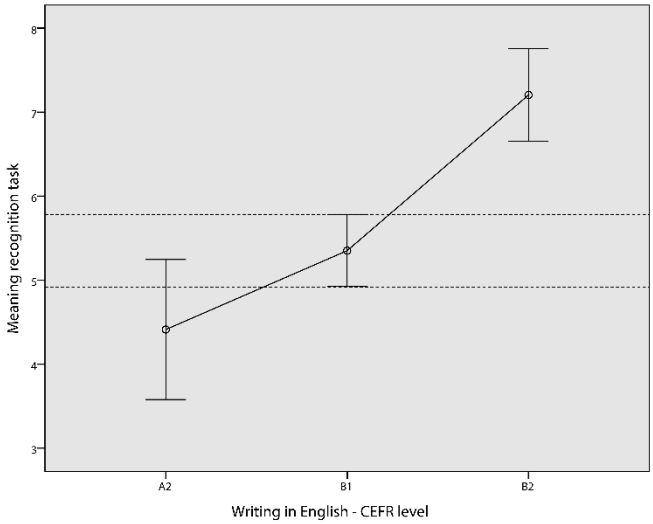


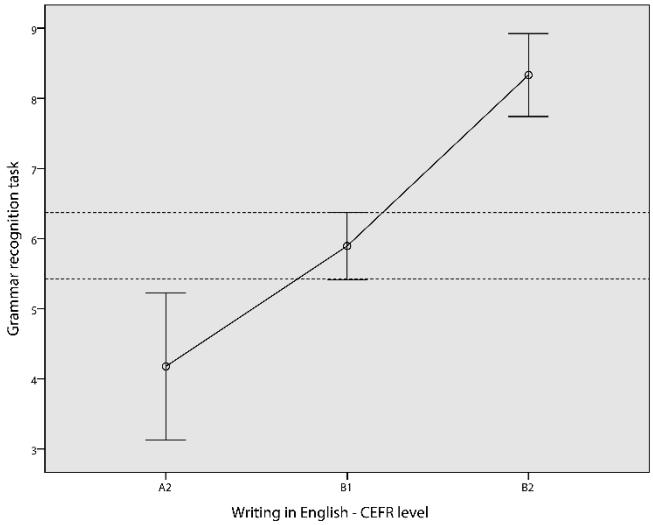


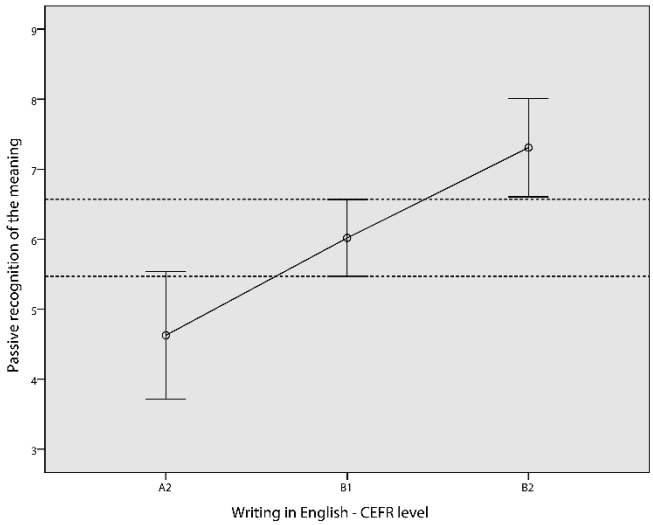


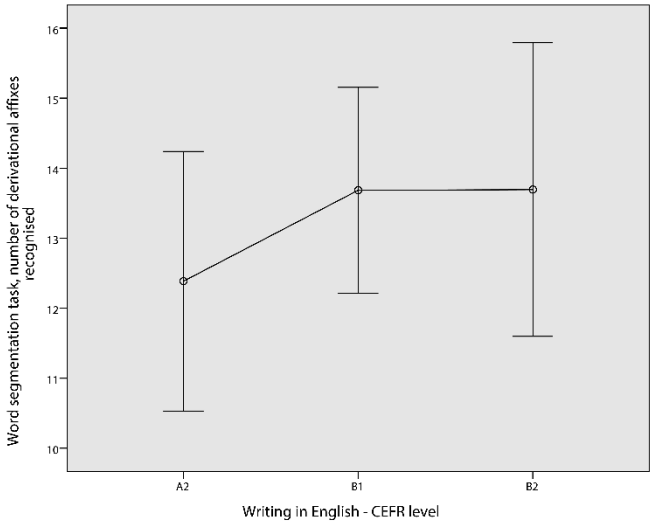

Prefix elicitation task
10
8
6
4
2
0
A2
B1
B2
Writing in English - CEFR level

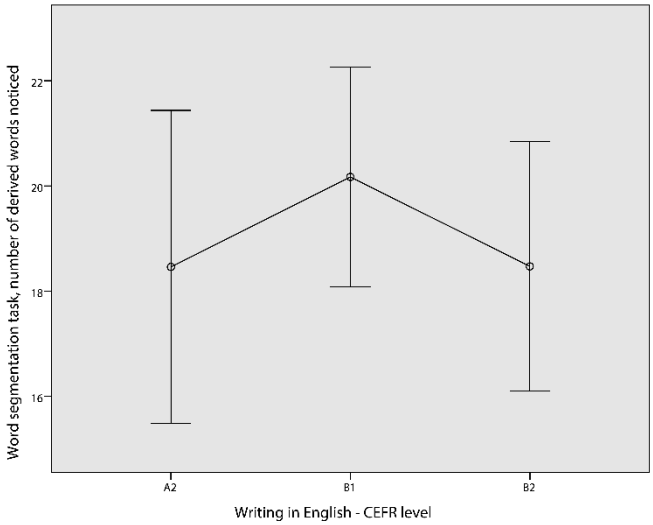

Non-word affix elicitation task
6
4
2
0
A2
B1
B2
Writing in English - CEFR level


Meaning recognition task
8
7
6
5
4
3
A2
B1
B2
Writing in English - CEFR level


Grammar recognition task
9
8
7
6
5
4
3
A2
B1
B2
Writing in English - CEFR level


Passive recognition of the meaning
9
8
7
6
5
4
3
A2
B1
B2
Writing in English - CEFR level





References

Akande, A. T. (2003). Acquisition of the inflectional morphemes by Nigeria learners of English. *Nordic Journal of African Studies*, *12*(3), 310-326.

Alanen, R., & Kalaja, P. (2010). *The emergence of L2 English questions across CEFR proficiency levels*. Paper presented at AAAL 2010, Atlanta, Georgia, USA.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. New York, NY: Continuum.

Alderson, J. C., Clapham, C., & Steel, D., 1997. Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research, 1*(2), 93–121.

Alegre, M., & Gordon, P. (1999). Rule-based versus associative processes in derivational morphology. *Brain and Language*, *68*, 347–354.

Bachman, L. & Cohen, A. (1998). *Interfaces between second language learning and language testing research.* Cambridge University Press.

Bartning, I., Martin, M. & Vedder, I. (Eds.) (2010) *Communicative proficiency and linguistic development: Intersections between SLA and language testing research.* EUROSLA Monograph series 1. European Second Language Association. Retrieved February 9, 2015, from http://eurosla.org/monographs/EM01/EM01home.html.

Bauer, L, & Nation, I. S. P. (1993). Word Families. *International Journal of Lexicography, 6*(4), 253–279.

Carlsen, C. (Ed.) (2013). Norsk profil. Det europeiske rammeverket spesifisert for norsk. Et første steg. Oslo, Novus forlag.

Chuenjundaeng, J. (2006). *An investigation of SUT students' receptive knowledge of English noun suffixes.* (MA thesis), Suranaree University of Technology.

Collins, B., & Nation, P. (2015). Testing receptive knowledge of derivational affixes. *Journal of Second Language Teaching & Research*, *4*(1), 6-23.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment* [electronic version]. Retrieved February 10, 2015, from http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf.

Extra, G., Spotti, M., & van Avermaet, P. (Eds.) (2009). *Language testing, migration and citizenship: Cross-national perspectives on integration regimes*. Continuum.

Finnish National Board of Education (2004). *National Core Curriculum for Upper Secondary Schools 2003* [Unofficial Translation]. Retrieved February 10, 2015, from http://www.oph.fi/download/47678_core_curricula_upper_secondary_education.pdf.

Friedline, B. E. (2011). *Challenges in the second language acquisition of derivational morphology: from theory to practice.* (Ph.D. dissertation), University of Pittsburgh.

Fulcher, G. (2003). Interface design in computer-based language testing. *Language Testing, 20*(4), 384–408.

Glaboniat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlag, L. (2005). *Profile Deutsch*. Langenscheidt.

Green, A. (2012). Language functions revisited. Theoretical and empirical bases for language construct definition across the ability range. Cambridge University Press.

Hayashi, Y., & Murphy, V. (2010). An investigation of morphological awareness in Japanese learners of English. *The Language Learning Journal, 39*(1), 105–120.

Huhta, A. (2012). Common European Framework of Reference. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell.

Hulstijn, J. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal, 91*, 663–667.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 36–55.

Leontjev, D. (2014). The effect of automated adaptive corrective feedback: L2 English questions. *APPLES: Journal of applied language studies, 8*(2), 43-66. Retrieved from http://apples.jyu.fi/ArticleFile/download/459.

Linacre, J. M. (2015). Facets computer program for many-facet Rasch measurement, version 3.71.4. Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2015). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com.

Mäntylä, K. & Huhta, A. (2013). Knowledge of word parts. In J. Milton. & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge*. Palgrave Macmillan, 45-59.

Martyniuk, W., & Noijons, J. (2007). *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe member states*. Strasbourg: Council of Europe. Retrieved February 9, 2015, from http://www.coe.int/t/dg4/linguistic/Source/Survey_CEFR_2007_EN.doc.

Milton, J., & Fitzpatrick, T. (Eds.)(2013). *The Dimensions of Word Knowledge*. Palgrave Macmillan.

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: an exploratory study. *System, 28*(2), 291–304.

Nagy, W. E., Diakidoy, I.-A. N., & Anderson, R. C. (1993). The Acquisition of Morphology: Learning the Contribution of Suffixes to the Meanings of Derivatives. *Journal of Literacy Research, 25*(2), 155–170.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Põhikooli riiklik õppekava õigusakt; Lisa 2 [Basic School National Curriculum Act: Annex 2] (2010). Pub. L. No. RT I 2010, 6, 22. Retrieved February 10, 2015, from https://www.riigiteataja.ee/aktilisa/1281/2201/0017/13275423.pdf.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77–89.

Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

Ringbom, H. (1990). On the relation between second language comprehension and production. In J. Tommola (Ed.), *Foreign language, comprehension and production: AFinLA yearbook* (pp. 139–148). Helsinki: Suomen soveltavan kielitieteen yhdistys.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281–317.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Macmillan.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*(01), 17–36.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*, 145–171.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361.

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language, 28*(6), 649–667.

Ullman, M.T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4*(02), 105–122. doi: 10.1017/S1366728901000220