

DOI: 10.2478/rjes-2014-0033

TESTING AND ASSESSING ENGLISH WORD-FORMATION

ANDREJ STOPAR

Faculty of Arts, University of Ljubljana

Abstract: *The paper discusses the testing/assessment of word-formation skills in threshold-level (B1) learners. To determine the suitability and use of productive gap-fill exercises at this level, an example of a word-formation exercise is addressed from the following perspectives: word frequency (corpus data), morphological complexity (morpheme types and word-formation processes), syntactic environments (typical sentence and/or phrase patterns), analysis of statistical data (facility values) and test-takers' errors. The text also addresses some implications of the findings for the instruction of lexis.*

Keywords: *assessment, CEFR, error analysis, testing, word-formation*

1. Introduction

Vocabulary is a basic element of language proficiency, yet it has not always been given the attention it deserves. The explicit instruction of vocabulary has been described as unfashionable (Nation 1990) since some believed that the acquisition of lexis occurs naturally, and consequently does not necessarily require explicit instruction (Moir and Nation 2008). Such views have been challenged by a number of authors (for instance, Alemi and Tayebi 2011; Nation 2001; Schmitt 2008; Schmitt and Zimmerman 2002), who claim that L2 instruction should take into consideration both intentional and incidental vocabulary learning.

With regard to derivation and word families, it has been established that learners' productive knowledge of the derivative forms of a word family is an essential part of vocabulary knowledge (Nation 1990, Schmitt and Meara 1997) and that gaining such knowledge does not seem to occur incidentally. Thus, Schmitt and Zimmerman (2002: 145) suggest that there may be "a need for more direct attention to the teaching of derivative forms." In fact, they even claim that learners essentially have to memorize these forms because they are so inconsistent in English.

This article builds on the above research by addressing various aspects of assessing and testing English word-formation and by examining a widely used word-formation task type: a productive gap-fill exercise that has the aim of assessing learners' productive knowledge of the derivative forms of a word family. An example of such an exercise is used to explore a set of factors that might facilitate or hinder L2 learners in their production of derivative (and, as the analysis shows, also inflectional) forms within a word family. The discussion is inspired by the wide-spread use of this task type in international examinations (for instance, FCE, CEA, CP), and the recent removal of a similar exercise from one of the nationally administered secondary-school-level examinations in Slovenia.

By analysing a productive gap-fill exercise and some common types of learners' errors, the paper aims to show how the data on word frequency (corpus data), morphological

complexity (morpheme types and word-formation processes), syntactic environments (typical sentence and/or phrase patterns) and statistical data (facility values) may be employed to improve the ways of assessing and testing L2 learners' skills related to derivation and inflection, to reflect on the approaches and materials used in teaching word-formation, and to facilitate both incidental and intentional learning of vocabulary (Schmitt 2008; Alemi and Tayebi 2011).

2. The research

The research is an investigation of learners' productive derivational (and inflectional) knowledge of seven words which were chosen as testing items as part of one of the Slovenian national examinations for professional secondary schools (the examination is usually referred to as the "vocational *matura*"). The research questions were as follows.

a. Do the morphological processes involved in the production of correct answers affect the facility values of test items?

b. Do various syntactic patterns affect facility values?

c. Is there a correlation between word frequency and facility values?

d. Do items placed at higher CEFR levels show lower facility values?

f. Do items with lower facility values exhibit a more extensive range of answers?

The study included 200 non-native learners of English, all of them students taking the national examination in English after completing their education at one of the Slovenian professional secondary schools. By the time the test-takers qualified to take the examination, they had had at least eight years of ESL instruction. The school-exit CEFR level (Council of Europe 2001), as defined in the English curriculum aims for Slovenian professional secondary schools, is B1.

The word-formation task analysed is based on a 300-word authentic English text on the topic of health. There are eight gaps in the text: the first one is provided as an example (*significant > significantly*), while for the rest of the gaps, the test-takers were instructed to use the word given in capitals to form a word that fits in the gap. The task was devised independently of this study by a committee of five item writers – all of whom are teachers of English, three of them teaching at professional secondary schools. It should also be noted that the task was devised when some of the online tools presented in subsequent sections were not available to item writers. Also, they had not received any specialist training with regards to CEFR levels. Their selection of the items was thus based predominantly on their teaching experience and their prior experience as task writers. No piloting scheme was in place for the examination, even though such practices are encouraged by experts in the field of testing (cf., for instance, Alderson, Clapham and Wall 1995). The word-formation task was part of a longer, 120-minute written examination that included an additional section on language and separate sections on reading and writing.

The seven prompt words in the task were as follows: *low*, *bleed*, *harm*, *health*, *different*, *participate*, and *normal*. The answers that the test-takers were expected to produce were: *lower*, *blood*, *harmful*, *healthy*, *difference*, *participants*, and *normally*.

A statistical analysis of the data provided information on the facility and the discrimination values for individual items. Also, word frequencies for both prompts and answers were examined (The British National Corpus 2007), as well as CEFR levels according to the Cambridge Dictionaries Online website (2013). Statistical analysis was conducted to investigate correlations between facility values, word frequencies and CEFR levels.

3. Results and discussion

Table 1: Facility and discrimination values (items are listed from least to most difficult)

	Prompt	Answer	Facility	Discrimination
Item 1	low	lower	0.64	0.78
Item 5	different	difference	0.60	0.82
Item 4	health	healthy	0.55	0.60
Item 2	bleed	blood	0.38	0.73
Item 3	harm	harmful	0.32	0.55
Item 7	normal	normally	0.31	0.58
Item 6	participate	participants	0.16	0.42
Average	-	-	0.42	0.64

Table 1

shows that facility and discrimination values fall within expected ranges. Item 6 (*participate* > *participants*) is the only one with a noticeably low facility value. The presented values also reveal that Items 2, 3 and 7 were the most challenging for the 200 test-takers. The average item difficulty of 0.42 indicates that the task was quite demanding. On the other hand, the discrimination indices are relatively high, suggesting that the task was a reliable indicator of the differences between weak and strong students.

3.1 Morphological Processes

Table 2: Morphological processes involved in word-formation

	Prompt	Answer	Types of Morphemes
Item 1	low	lower	stem + comparative derivational affix; zero derivation (adjective to verb)
Item 2	bleed	blood	stem (vowel change; OE)
Item 3	harm	harmful	stem + derivational affix
Item 4	health	healthy	stem + derivational affix
Item 5	different	difference	stem + derivational affix
Item 6	participate	participants	stem + derivational affix; inflectional affix
Item 7	normal	normally	stem + derivational affix

As

indicated in Table 2, most of the items require the formation of derivatives by suffixation. The exceptions are Items 1, 2 and 6. For Item 1, the etymology of the verb *lower* reveals that it comes from the comparative of *low*, so to get the correct answer, a two-step process is required from the test-takers: the formation of the comparative with the suffix *-er*, followed by zero derivation (Partridge 1983: 352). To continue, the move from *bleed* to *blood* in Item 2 does not involve derivation – in fact, the verb was derived from *blōd* in the Old English period (Partridge 1983: 49). And, finally, the correct answer *participants* involves both derivation and inflection since it consists of the derivational morpheme *-ant* and the inflectional plural morpheme *-s*.

Items 3, 4, 5 and 7 all involve derivation. The adjectival derivational suffixes *-ful* and *-y* in Items 3 and 4 are among the most common such suffixes in English (cf. for instance Biber et al. 2004: 531; Adams 2001: 36). Similarly, the nominal ending *-ence* in Item 5 is listed among the most common derivational morphemes in nouns (Biber et al. 2004: 321), and the adverbial suffix *-ly* in Item 7 as one of the most common English adverbial endings, especially in news articles (Biber et al. 2004: 540; Adams 2001: 38-39).

The evidence suggests that the complexity of morphological processes required in the formation of answers does not necessarily influence item difficulty. This can be exemplified by a brief comparison of Items 6 and 7. Item 6, which involves both derivation and inflection,

is the most difficult item in the test (cf. Table 1) and thus supports the prediction of complex morphological processes affecting item facility values. Yet, Item 7, which involves straight-forward affixation with a very frequent suffix, is, contrary to expectation, the second most difficult item in the test.

3.2 Syntactic Patterns

Table 3: Test items in their syntactic environments

	Context
Item 1	... make a big enough change to LOWER the risk of ...
Item 2	... decreased their total BLOOD cholesterol by ...
Item 3	... HDL is protective rather than HARMFUL.
Item 4	... the two kinds of cholesterol are considered to be HEALTHY.
Item 5	... to see a DIFFERENCE between the two doses ...
Item 6	At the start of the study PARTICIPANTS ate an average American diet ...
Item 7	... antioxidant NORMALLY found in dark leafy vegetables.

Table 3

provides context to present the syntactic environments where the gaps were placed. The items (correct answers) are capitalized.

In Item 1 the pattern is “verb + nominal phrase + to-infinitival clause”. According to Biber et al. (2004: 698-9), this is the second most-common pattern for infinitive clauses; it is most typical in written registers, in particular in news articles. Corpus data show that the frequency of such structures is 500-1000 per 1 million words.

Item 2 is a nominal premodifier in a complex nominal phrase headed by a common noun. Such structures are not rare, especially in written registers (Biber et al. 2004: 581). Nouns as premodifiers are not as common as adjectives; however, they still account for 30-40 percent of all premodifiers in news and academic prose.

Items 3 and 4 contain linking structures introduced by the very common copula *be*. The copula in Item 3 is followed by a frequent, albeit complex adjectival complement, while in Item 4 it is embedded in a – much rarer – complement clause (Biber et al. 2004: 446).

Item 5, is placed in-between an indefinite article and a preposition, which clearly signals that a nominal element is required to fill the gap.

In Item 6 we see that the missing word is the subject of the sentence. However, the gap is preceded by a complex adverbial. According to Biber et al. (2004: 772) in less than 15 percent of circumstance adverbials appear in the sentence-initial position. Furthermore, that the last word of the adverbial is a noun and the six-word adverbial structure is not separated from the rest of the sentence by a comma does not make this structure any easier to process.

Finally, Item 7 is an adverb modifying a verb in a restricted relative clause; this is a frequent position for adverbs.

The above indicates that Items 3, 4 and 6 might present a challenge for test-takers due to the syntactic complexity of the structures that they are part of. However, the facility values in Table 1 do not support such a prediction fully. Whereas Items 3 and 6 are indeed among the more demanding ones, Item 4 is has a high facility value.

3.3 Word Frequency, CEFR Levels and Facility Values

As facility values in Table 1 indicate, Items 2, 3, 6 and 7 were the most challenging for test-takers. We have shown that the difficulty of the items can be only partly accounted for by the morphological and syntactic processing involved in the production of answers. This

section will attempt to provide more conclusive observations by shedding some light on the impact of word frequency and CEFR levels on item difficulty.

The British National Corpus (2007) and Cambridge Dictionaries Online (2013) were used to determine word frequencies and CEFR levels of the test items. Table 3 below repeats the information on facility values and adds the information on word frequency and CEFR level.

Table 3: Facility values, word frequency and CEFR levels

	Difficulty	Answer Frequency	Prompt CEFR Level	Answer CEFR Level
Item 1	0.64	12263	3 (B1)	4 (B2)
Item 5	0.60	11167	1 (A1)	2 (A2)
Item 4	0.55	3528	2 (A2)	2 (A2)
Item 2	0.38	9778	3 (B1)	2 (A2)
Item 3	0.32	813	4 (B2)	4 (B2)
Item 7	0.31	8130	2 (A2)	3 (B1)
Item 6	0.16	2227	4 (B2)	5 (C1)
Average	0.42	-	2.71	3.14
Correlation	-	0.63	0.62	0.52

With regard to word frequency, it can be observed that the word with the highest frequency is in fact the word with the highest facility value, in other words, the easiest item (cf. Item 1 (*low* > *lower*)). Similarly, the words with lower frequency counts tend to be more difficult as test items (for instance, Item 6 (*participate* > *participants*)). The moderately high correlation between the frequency counts for answers and the facility values (0.63) suggests that word frequency can serve as a relatively reliable predictor of item difficulty. The exceptions are Item 4 (*health* > *healthy*) and Item 7 (*normal* > *normally*).

The data on the CEFR levels show that the prompts are placed slightly below the B1 level, while the answers, on average, are above this level. What is also noticeable is that only one of the answers is on the B1 level, while the rest of them are either lower or higher on the CEFR scale. Especially noticeable are three items at the A2 level and one item at the C1 level. The prompts are more balanced in this regard. Furthermore, our statistical analysis shows that the correlations between CEFR levels and item difficulty are slightly lower than those for frequency. However, they remain relevant – for more than half of the items, the CEFR word level seems to be a relevant predictor of item difficulty. Additionally, the analysis also suggests that the CEFR level of prompts is a better predictor of difficulty than the level of the answers: the correlations are 0.62 and 0.52, respectively.

The frequencies of Items 3 (*harm* > *harmful*), 6 (*participate* > *participants*) and 7 (*normal* > *normally*) are comparatively low; they are also placed at higher CEFR levels. The facility values confirm the relevance of these factors. On the other hand, Item 4 (*health* > *healthy*) did not represent a substantial challenge for the test-takers despite its relatively low frequency count. Its facility value can be explained by its low CEFR placement – the item is simply not demanding enough for the B1 test-takers.

We can conclude that word frequency and CEFR levels exhibit a moderate correlation with facility values.

3.4 Range of Answers and Error Analysis

The analysis shows that six of the seven items elicited between 20 and 26 different answers each. For instance, among the 20 different word forms provided for Item 1 were *lowed*, *lowly*, *downlowed*, *lover*, *significant* and so on. One of the items, however, stands out with regards to the variety of answers – the test-takers were most creative when dealing with Item 6 (*participate* > *participants*): they provided 51 different word forms as their answers.

Not surprisingly, the correlation between the number of answers and item difficulty is moderately high at 0.63.

The correct answers were typically the most numerous ones. The only exception is Item 7 (*normal* > *normally*) since the most common answer for this item was *normaly* (spelled with a single *l*).

Table 4: Types of answers

	Types of Answers	Percent	Example
Group 1	Correct answers	42	<i>lower</i>
Group 2	Wrong answers but correct part of speech	22	<i>lowed</i>
Group 3	Wrong answers and wrong part of speech	19	<i>lowest</i>
Group 4	Wrong spelling	14	<i>lover</i>
Group 5	Answers belong to some other word family	1	<i>reduce</i>
Group 6	No answer	2	-

Table 4

shows that among the wrong answers the most numerous are those that suggest that the test-takers were aware what part of speech was required, but they did not know what suffix to use to form the correct answer (cf. Group 2; 22 percent). To exemplify, the analysis of wrong answers for Item 3 shows that 41 percent of students provided wrong answers that still included typical adjectival endings such as *-y* (*harmly*). This finding supports the claim by Schmitt and Zimmerman (2002) about the English derivational system being deceptively regular. Furthermore, the test-takers' ability to discern from the context the syntactic and morphological requirements while still providing a wrong answer indirectly shows that although there is a correlation between suffix knowledge and total vocabulary size (Schmitt and Meara 1997), the correlation is not necessarily high.

A group of test-takers that is almost numerically equal to the previous group includes answers that suggest their inability to identify even the part of speech (cf. Group 3; 19 percent). It is safe to assume that this is the group showing the weakest proficiency with regards to word-formation, and, indirectly, syntax and morphology.

Also numerous is Group 4, which includes the test-takers who only had problems with spelling (14 percent).

If the above numbers are considered from the perspective of English sentence structure and morphology, the result for the test is not necessarily as poor as the average facility value of 0.42 might suggest (cf. Table 1). In fact, a closer look at groups 1, 2 and 4 in Table 4 reveals that 78 percent of test-takers knew what part of speech to use. Furthermore, since our students are often described as digital natives (Bennett, Maton and Kervin 2008), who have been immersed in technology since they were born and regularly use tools such as spell-checkers and autocorrect, and since the test in question was a paper-to-pen test, perhaps the causes of the spelling mistakes themselves should be reconsidered. Specifically, are answers such as *normaly* in Group 4 not higher on the cline of correctness than answers such as *normales*? Admittedly, both are incorrect; yet disregarding the wrong spelling for Item 7 (*normal* > *normally*) would substantially alter its facility value: the change from 0.31 to 0.87 would make it one of the easiest items in the test.

4. Conclusion

The analysis of a productive gap-fill exercise and some typical errors observed in the group of 200 test-takers revealed a number of factors that influence test item difficulty. The research questions focussed on aspects of word-formation tasks related to morphological processes, syntactic environments, word frequency, CEFR levels and range of answers. The findings suggest that morphological processes involved in the formation of correct answers and the syntactic environments in which the items are placed are not reliable indicators of

item difficulty. Naturally, they should be taken into account when testing, assessing or teaching word families; however, the factors that allow test writers or teachers to predict the suitability of a word-formation task more accurately seem to be word frequency and CEFR levels. Relying on corpora and vocabulary profiles should also help reduce the range of test-takers' answers, since a link has been established between the difficulty of the item and the number of different (wrong) answers provided by test-takers. It can be concluded that when devising or analysing a word-formation test, all of the above factors should be considered – but with an important caveat: since the study is based on the analysis of only seven test items, further research is required to obtain more conclusive findings.

The discussion has also brought to light a number of other relevant issues that should affect the testing, assessing and teaching of vocabulary. Firstly, with regards to good practices in language testing, the need for piloting and pretesting should be mentioned – such procedures enable test writers to remove overly demanding items before administering the test (Alderson, Clapham and Wall 1995). Secondly, the facility values reveal only partial information on the test-takers' vocabulary knowledge: a more thorough error analysis sheds some light on the types of mistakes they make. For instance, the analysis suggests that the majority of test-takers have a thorough knowledge of syntax and morphology, i.e. they are able to identify the correct part of speech and select a suffix that is typical for the identified word class. Another relevant observation is that spelling errors have a substantial effect on test results, which should be taken into account when focussing on word families in the classroom.

As mentioned above, studies have shown that instructed vocabulary learning is important (Alemi and Tayebi 2011, Moir and Nation 2008; Nation 1990; Schmitt 2008; Schmitt and Meara 1997; Schmitt and Zimmerman 2002), so the role of the teacher is not only to facilitate incidental learning but also to identify students' needs and provide opportunities for intentional learning. With regards to productive gap-fill exercises, teachers should learn from the mistakes of their students and, for instance, rely on intentional learning when explaining to their students that a certain suffix is correct for the desired word class but not for every word belonging to that class. Only in this way will the learners' knowledge of English word-formation steadily progress from *wonderfull* to wonderful.

References

- Adams, Valerie. 2001. *Complex Words in English*. Harlow: Pearson Education Limited.
- Alderson, J. Charles, Caroline Clapham and Dianne Wall. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alemi, Minoo and Alireza Tayebi. 2011. "The Influence of Incidental and Intentional Vocabulary Acquisition and Vocabulary Strategy Use on Learning L2 Vocabularies." in *Journal of Language Teaching and Research* 2 (1), pp. 82-98.
- Bennett, Sue, Karl Maton and Lisa Kervin. 2008. "The 'digital natives' debate: A critical review of the evidence" in *British Journal of Educational Technology* 39 (5), pp. 775-786.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 2004. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Cambridge Dictionaries Online. 2013. Available: <http://dictionary.cambridge.org/> [Accessed 2013, March 14].
- Council of Europe. 2001. *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press
- Moir, Jo and Paul Nation. 2008. "Vocabulary and good language learners" in *Lessons from Good Language Learners*. Carol Griffiths (ed.). Cambridge: Cambridge University Press, pp. 159-173.
- Nation, I. S. P. 1990. *Teaching and learning vocabulary*. New York: Newbury House.
- Partridge, Eric. 1983. *Origins: A Short Etymological Dictionary of Modern English*. New York: Greenwich House.
- Schmitt, Norbert. 2008. "Instructed second language vocabulary learning." in *Language Teaching Research* 12 (3), pp. 329-363.
- Schmitt, Norbert and Paul Meara (1997). "Researching vocabulary through a word knowledge framework (Word

associations and verbal suffixes)” in *Studies in Second Language Acquisition* 1, pp. 17-36.

Schmitt, Norbert and Cheryl Boyd Zimmerman. 2002. “Derivative Word Forms: What Do Learners Know?” in *TESOL QUARTERLY* 36 (2), Summer, pp. 145-171.

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available: <http://www.natcorp.ox.ac.uk/> [Accessed 2013, March 14].

Notes on the author

Andrej STOPAR, PhD, is a lecturer at the English Department at the Faculty of Arts, University of Ljubljana, Slovenia. His research interests are primarily focussed on (generative) syntax, discourse analysis, translation and language testing. The courses he has taught include Language in Use, English Grammar for Teachers, Translation (Slovenian to English and English to Slovenian) and Slovenian Morphology.