Article

LANGUAGE TESTING

Understanding test-takers' perceptions of difficulty in EAP vocabulary tests: The role of experiential factors

Language Testing
2017, Vol. 34(3) 413–433

DOI: 10.1177/0265532216673399
journals.sagepub.com/home/ltj
SAGE

Nesrin Oruç Ertürk and Simon E. Mumford
Izmir University of Economics, Turkey

Abstract

This study, conducted by two researchers who were also multiple-choice question (MCQ) test item writers at a private English-medium university in an English as a foreign language (EFL) context, was designed to shed light on the factors that influence test-takers' perceptions of difficulty in English for academic purposes (EAP) vocabulary, with the aim of improving test writers' judgments on difficulty. The research consisted of a survey of 588 test-takers, followed by a focus group interview, aimed at investigating the relative influences of test-taker factors and word factors on difficulty perceptions. Results reveal a complex interaction of factors influencing perceived difficulty dominated by the educational, and particularly, the social context. Factors traditionally associated with vocabulary difficulty, such as abstractness and word length, appeared to have little influence. The researchers concluded that rather than basing their intuitions regarding vocabulary difficulty on language-lesson input or surface features of words, EAP vocabulary test writers need a clear understanding of test-takers' difficulty perceptions, and how these emerge from interactions between academic, social and linguistic factors. As a basis for EAP vocabulary item writer training, four main implications are drawn, related to test-takers' social and educational background, field of study, the features of academic words, and the test itself.

Keywords

EAP testing, EAP vocabulary tests, test-taker factors, test-takers' perceptions

Corresponding author:
Nesrin Oruç Ertürk, School of Foreign Languages, Izmir University of Economics, Sakarya Caddesi No: 156, Balçova, Izmir, 35330, Turkey.
Email: nesrin.oruc@ieu.edu.tr

In the process of writing vocabulary items, test developers gauge and attempt to manipulate the level of item difficulty to create items that will serve their intended purpose. The success of item writing therefore depends in part on the item writers' judgments of vocabulary difficulty, but evidence suggests that test writers are not particularly good at predicting difficulty (Falchikov, 2005, p. 34). In vocabulary testing, the challenge of predicting item difficulty is partly owing to the fact that a word's actual difficulty for a test-taker results from complex interactions between the word's frequency and its formal features such as length and grammatical class (Milton, 2009, p. 42). Formal features can be calculated by test developers in a formulaic manner, but word frequency is less straightforward. Seemingly objective estimates of word frequency are based on particular corpora of texts and therefore need to be interpreted relative to those collections of texts. Moreover, especially for low frequency words, experiential frequency (i.e. the frequency of actual encounters by individual learners) may be a better guide to difficulty than published frequency lists (Gernsbacher, 1984).

Estimating the likely number and strengths of encounters test-takers have had with individual words is more or less feasible depending on the context and purpose of the test. For example, an achievement test written by a teacher who has taught the test-takers would be better able to estimate students' exposure to certain words than an item writer's proficiency test given to students across diverse educational settings. Nevertheless, even when the test developers know the students and curriculum, item difficulty results from a combination of word and learner factors. For this reason, Van de Watering and van der Rijt (2006) emphasize the need to understand difficulty from the test-takers' point of view, but research on test-takers has tended to focus on issues such as washback, reactions to tests, anxiety, and exam strategies, rather than perceptions of the tested knowledge. For example, Bonaccio and Reeve's (2010) investigation of test-taker perceptions of anxiety relates to test-takers (e.g. level of preparedness) and the test (e.g. level of cognitive difficulty), rather than a direct focus on tested knowledge.

In order to yield insights for developers of vocabulary tests, the current research investigates the complex interaction of factors that affect test-takers' perceptions of vocabulary difficulty. Unlike previous research that investigated test-takers' perceptions of particular test formats or the relationship between test-taker characteristics and their perceptions (Bradshaw, 1990; Brown, 1993; Huhta, Kalaja, & Pitkanen-Huhta, 2006; and Xiao & Carless, 2013), this study investigates test-takers' perceptions of what makes the test content difficult. Like the study by Kim and Elder (2014) that focused on test-taker perceptions of the tested language in a test of aviation English in Korea, this study gathers data about perceived sources of difficulty in the vocabulary tested. The approach to data collection is informed by theory and research on the factors affecting vocabulary learning, including test-takers' prior experience. The need for language testers to take into account the test-takers' wider knowledge and the educational background was recognized by Shohamy (2001, p. 387), but has not been fully studied in vocabulary testing. Test-takers' previous experience is included in the following discussion of factors affecting word difficulty.

Factors affecting word difficulty

The study of test-takers' perception is based on a survey and interview that elicit test-takers' perceptions about two theorized dimensions of vocabulary difficulty: frequency

of exposure and word factors (e.g. length, form). These two factors interact with each other and with other factors including the learner and the test methods.

Frequency of exposure

Frequency of exposure to English vocabulary in an English-medium university in an EFL context depends to some extent on a student's experience in their academic area of study. Technical texts demand a high level of specific vocabulary knowledge (Waring & Nation, 2004), but, as Wesche and Paribakht (2009) point out, the texts themselves are a source of learning. L2 texts, even without language instruction, can contribute to vocabulary learning, albeit slowly, by providing a source of material from which students can learn through lexical inferencing, which is the key process in incidental learning (Hulstijn, 2003).

In addition to academic texts as a source of exposure, O'Sullivan (2000, described in Elliott, 2013, p. 37) noted the importance of a broader set of experiential factors. In an EFL university context, previous and current learning experiences are of key importance. Experiential factors affect learners' exposure to words in both academic and non-academic contexts. The former concerns field of study; not all academic vocabulary is equally relevant to specific disciplines. The academic word list (AWL) (Coxhead, 2000), which has been widely used by materials developers as a basis for identifying frequent academic vocabulary, favors disciplines such as economics and law (Hyland & Tse, 2007). Test-takers' perceptions of the difficulty of these words would therefore be expected to be affected by their exposure to texts in these subject-specific materials. Exposure to academic language in the non-academic context will vary according to each learner's level and type of participation in English-medium social environments, which Corson (1997, p. 680) argues has a major role in academic vocabulary acquisition.

Word factors

Frequency of exposure effects interact with factors relating to knowledge of form and meaning (Milton, 2009), including "the spoken and written form, morphological knowledge, knowledge of the word meaning, collocational and grammatical knowledge, connotive and associative knowledge" (Laufer & Goldstein, 2004, p. 400). This study considers those aspects that would be understood by tertiary-level test-takers, namely, pronunciation, abstractness, multiple meanings, grammatical form, and affixes, avoiding specialist terminology such as collocation and connotation.

Semantic field. Academic vocabulary excludes basic vocabulary, and consists of words common in a wide range of academic texts, and which are uncommon in non-academic texts, referred to as sub-technical vocabulary (Scarcella & Zimmerman, 2005). The AWL is considered the definitive list for teaching and testing in the current context, in line with its intended purpose (Coxhead, 2000). In the current research, the term *academic*/*EAP vocabulary* is synonymous with the AWL. This vocabulary tends to be learnt incidentally; according to Ellis and Shintani (2014), owing to its size, it is mainly acquired through subject-specific study rather than formal language instruction.

Pronunciation. The Phonological Loop Theory emphasizes the role of the vocalization of items in allowing items to pass into the longer term memory (Baddeley, Gathercole, & Papagno, 1998). Recent research has emphasized the importance of the phonological short term memory, and therefore, pronunciation, for L2 vocabulary acquisition (for a review, see Martin & Ellis, 2012). Pronunciation also has a social aspect; according to Hansen Edwards (2008, p. 253), L2 pronunciation is affected by learners' social proximity to other users of the language; the closer the ties, the more accurate the pronunciation, highlighting the benefits for native or fluent L2 speakers in the social environment.

Word form, affixes, and polysemy. In this study, grammatical form, affixation and polysemy are considered as being closely related. Difficulty may be affected by grammatical form; in L1 learning, nouns are acquired earlier, perhaps owing to greater imageability (i.e. ease of visualization), which is linked to meaningfulness (Ellis & Beaton, 1993, p. 565). For ESL students, nouns and verbs were found to be easier than adverbs and adjectives (Schmitt & Zimmerman, 2002). In the case of academic vocabulary, even when one or more forms are known, knowledge of all forms is rare (Scarcella & Zimmerman, 2005).

English morphology, particularly affixation, can cause problems because of irregularity and the effect of deceptive morphological structure (i.e. the tendency of stem forms to resemble affixes) (Schmitt & Zimmerman, 2002). These difficulties have implications for the AWL, as different forms of the same word may have different meanings, not all of which will be relevant to particular disciplines. Wang Tzu and Nation (2004) give 10 separate senses for *neutral*, some of which relate to different content areas. Similarly, Hyland and Tse (2007) noted semantic variation between different forms of the same word, citing the word *process*, which relates to different fields depending on whether it is a verb or a noun. In addition, as Paquot (2010) notes, academic words may simultaneously have both general meanings and "extended meanings in specific disciplines." So although polysemy is a source of difficulty for all learners, it is a greater problem for academic learners faced with a multitude of words with multiple meanings, not all of which are relevant to their field.

Abstractness/concreteness. According to Bachman (2003, p. 135), abstract representations "are primarily symbolic or linguistic, whereas concrete information is capable of representation in other than linguistic modes." Ellis and Beaton (1993) found imageability to be strongly associated with learnability; De Groot and Keijzer (2000) consider imageability to be synonymous with concreteness, and highlight difficulties caused by the lack of "context availability" (i.e. the inability to think of contexts for abstract words). Many AWL words are abstract, thus increasing propositional density, and potentially, difficulty for test-takers (Paquot, 2010). Nevertheless, the direct relationship between abstractness and difficulty has been challenged by Context Availability Theory, which proposes that while understanding concrete words involves sensory and motor functions, abstract words are more connected to the social and emotional environment (Connell & Lynott, 2012). This suggests that the difficulty associated with abstractness is not an intrinsic quality of words, and may vary from learner to learner.

Interactions and other factors

The complexity of the interactions among factors is demonstrated by Milton (2009), who gives the illustration of cognates; cognateness can overcome the difficulty associated with infrequency, but cognates are likely to be longer words, so the concept that length necessarily increases difficulty is undermined. Moreover, any individual learner may have more or less experience with a particular cognate because of his or her first language and field of study. Speakers of Turkish, for example, may perceive English vocabulary as easier when they know the cognate form in Turkish. Turkish has a tendency to adopt English words (Dogancay-Aktuna & Kiziltepe, 2005). Even though cognates may have an important role in reducing difficulty, because corresponding words are represented as a single item in the bilingual memory (De Groot & Keijzer, 2000, p. 33), their effect on difficulty is rather unpredictable; for example, a particular cognate may not be recognizable to every learner (Milton, 2009, p. 36).

Such complex interactions make it difficult to create a single model of difficulty, and emphasize the need to understand the interaction between the test-taker and the tested knowledge in specific contexts. In addition, Bachman (2003, p. 164) has pointed out the effect of test method facets on test-takers' performance in language tests. He describes test method facets that affect the perceived and actual difficulty of test items (Bachman, 2003, p. 115). Even the familiar multiple-choice questions (MCQs) used in the current context can be influenced by factors such as format and length. Furthermore, effective MCQs are difficult to write, and distractors can have an excessive influence on the test outcome (Read, 2000).

The study

In view of the multiple interacting factors that may come into play in the actual difficulty of vocabulary test items, test item writers cannot rely on a single objective source of information such as its frequency, morphological characteristics or status as a cognate to judge difficulty during item writing. Instead, intuition needs to be developed through a solid understanding of how the interacting factors affect difficulty. An essential source of information for developing intuition is test-takers' perceptions about the sources of difficulty.

Research context and site

The study was conducted in 2013 at an English-medium private university in Turkey. Except for a minority who pass an exemption exam, all students attend the year-long university English Preparatory Program. The program starts with a general-purpose course and progresses with an increasing EAP emphasis in the second semester, but the course is a general academic-purpose one, rather than a subject-specific one.

Following their attendance at the Preparatory School, first-year students take a compulsory English academic skills course, which is the context of this study. Coded as ENG101 (Fall Semester) and ENG102 (Spring Semester), this faculty-specific program provides academic skills and vocabulary support. The only texts are in-house

faculty-specific course books, standardized in terms of objectives and learning tasks. At the time of this research, the university consisted of five faculties (Engineering and Computer Sciences, Business Administration, Fine Arts and Design, Arts and Sciences, and Communication) and one school (Applied Management Sciences, Culinary Arts and Management).

The vocabulary learning and testing context

ENG 101/102 exams test the retention of AWL words presented in the course book's texts. Words are highlighted in the reading texts, and reinforced by multiple-choice or matching exercises. The course book also contains an appendix with dictionary references for these words, including forms and meanings. Approximately 70 words are presented in each of the two 15-week semesters (Appendix A). As only a small portion of the four-hours-per-week course is allocated for vocabulary teaching, detailed discussion of collocation patterns or connotations is impossible. Vocabulary sections are worth 15% of the mid-term and final exams; however, vocabulary is also emphasized in the grading criteria for written and spoken production, including an assessed oral presentation and an essay; therefore, vocabulary is a significant component of the course.

Test items are regularly reviewed by the Freshmen Testing Unit, of which both researchers were members. Based on our reviews of test items produced by the item writers, we recognized the limitations in the intuitions that were informing the test-writing practices. Without the test-taker perspective, test writers' view of difficulty depends on their intuitions based on their classroom experience, which reflects only a small part of learners' overall knowledge. The research aimed to provide a basis for richer insight that could inform the test-writing process. The two research questions are as follows:

1. To what extent do the different test-taker experience factors contribute to the test takers' perception of difficulty in the tests?
2. To what extent do the different word factors contribute to the test takers' perception of difficulty in the tests?

Methodology

The methodology relied primarily on quantitative data obtained from a survey developed especially for this study to assess test-takers' perception about vocabulary difficulty. The interpretations are based on triangulation of these quantitative data with qualitative data from a focus group interview (Anderson, Bachman, Perkins, & Cohen, 1991).

Participants

Participants were students at an English-medium university in Turkey. At the time of the study in 2013, they were taking the ENG102 (spring semester) course, having reached upper-intermediate (CEF B2) level, an IELTS equivalent of 5.5. All shared the same L1, Turkish.

Table 1. Participant information (N=588).

Participant characteristic	Response	n
Age	18	16
	19	90
	20	236
	21+	246
Gender	Male	345
	Female	243
Preparatory School attendance	Yes	538
	No	50
Length of English education (years)	0–1	39
	2–3	148
	4–5	47
	+5	354
Faculty[a]	Faculty of Business Administration;	104
	Faculty of Fine Arts and Design	112
	Faculty of Engineering and Computer	102
	Faculty of Arts and Sciences	112
	Faculty of Communication	110
	ulinary Arts and Management	48

The participants were 588 of a total of 1,392 ENG 102 students, who were chosen to include around 100 students from each faculty, and all 55 students in Culinary Arts and Management. Ages ranged from 18 to above 20. They were assured of anonymity and the right not to participate. No incentives for participation were offered. Table 1 gives the participant details.

The instrument

The first part of the survey elicited personal information, such as age, length of study, and faculty (Appendix B). The main section focused on the perceptions of difficulty in acquiring academic vocabulary caused by various test-taker and word factors. The vocabulary difficulty perception items were developed using information from the literature and the researchers' own expertise. In order to maximize the construct validity of the instrument, the target construct was conceptualized carefully by the researchers, ensuring that they both had the same understanding of the construct "perception of vocabulary difficulty." The construct was defined as the combination of factors that second language users perceive as contributing to their difficulty in correctly responding to vocabulary items on English language tests; there are two components to this construct, test-taker factors and word factors, each of which is tested with different items. Moreover, the researchers paid careful attention to the wording of questionnaire items (Clark & Watson, 1995). The construct validity of the survey was investigated through the use of factor analysis to identify the extent to which our survey captured the theoretical construct,

namely perception of vocabulary difficulty. The results revealed two factors explaining a total of 52.50% of the variance for the entire set of variables. Factor 1 was labeled "test-taker factors." Items 1–4 and 14 measured test-taker factors relevant to their exposure to vocabulary (e.g. "Attending the Preparatory School has made it easier for me to learn academic words"), and it explained 29.953% of the variance. The second factor derived was labeled "word factors" and it was labeled as such because of the high loadings by factors such as "Affixes make academic words harder to learn." Items 5–13 related to word factors. The variance explained by this factor was 22.55%.

Note that item 5 is considered a word factor, as it relates to the general frequency of words in texts, a well-established property of words in vocabulary research, and that item 13 refers to the formality of the texts in which the word is found, in parallel with item 11, which refers to the level of abstractness of the context. Familiar terms such as *affixation* (a distinctive feature of L1), *abstractness* and *multiple meanings* were included, whereas potentially confusing terminology, such as *collocation* and *connotation*, was avoided.

For items 1–14, a multi-item (five-item) scale (Classon & Dormody, 1999) was considered the most suitable for the age range. A Likert-type response scale ranged from 1 (strongly disagree) to 5 (strongly agree). The researchers conducted a pilot study to test the reliability of the scale, as detailed below. The survey was written in English, translated into the participants' L1, and then checked by a faculty member from the Translation Department.

The pilot study

A pilot study was conducted with the researchers' own classes from the Faculties of Business Administration, and Fine Arts and Design, totaling 59 students, in line with Baker's (1994) recommendation that 10–20% of the total sample size is acceptable for this purpose. The pilot study resolved issues related to the survey instructions and wording, and the reliability of the scale, and also helped to determine efficacious statistical and analytical processes.

A Cronbach α reliability coefficient of .86 was found for the pilot study, representing a good level of reliability for Likert-type scales according to George and Mallery (2003, p. 231), whose reliability standards are cited in the applied linguistics literature (e.g. Liu, Chang, Yang, & Sun, 2011). Therefore, only minor changes and clarifications were needed. The pilot study participants were excluded from the actual study.

Procedure

After being informed of the purpose of the study, 13 ENG102 teachers administered the surveys to their classes. Based on the pilot study, 20 minutes were allowed; in case of queries, teachers advised participants to choose what they felt was the most appropriate answer.

Focus group interviews

In July 2013, a 50-minute focus group interview was conducted, designed to explore themes emerging from the survey. For the focus group interview, 10 students (five female, five male) were chosen using purposive sampling to ensure representation across

all faculties. A guide (Appendix C) based on theory and the researchers' experience, was used for the semi-structured interviews (Harrell & Bradley, 2009). The interview was conducted by one researcher in the students' native language, to decrease anxiety levels, and also because only one shared the participants' L1. It was explained that anonymity would be guaranteed, and that pseudonyms would be used when citing participant comments. The interview was recorded, transcribed and translated, and comments were selected for inclusion in the discussion of results, based on their representativeness and potential to shed light on the survey results.

Results

This section provides survey results addressing each research question, and it adds to the interpretation with selected comments from the interview to shed light on the survey's results.

Test-taker factors

The first question was about the extent of the different test-taker factors' contribution to their perception of vocabulary difficulty in the tests. Table 2 shows the results for items 1–4 and item 14 relating to the Preparatory School, the field of study, the non-academic context, the L1, and the time for learning. Among these, the greatest influence on ease of learning was reported as encountering words outside the educational context, with 340 of 588 participants strongly agreeing (mean = 4.30). The role of the L1 (mean = 4.13) and the field of study (mean = 3.77) were also perceived as influential.

The Preparatory School (item 1) was perceived as being much less influential than items referring to frequency of encounter inside and outside the university (items 2–4). As a preparation for academic study, the School was expected to be influential; however, although more than 90% of respondents had attended (Table 1), their response to item 4 indicated that most respondents disagreed that attending the School had made learning academic vocabulary easier. One participant, Nurşah (22, Psychology), reported that the School was useful for general English, but "not at all for academic English," implying that a year is an insufficient amount of time in which to acquire the necessary language. This view was supported by Helin (19, Software Engineering), citing her own lack of language education, and showing an awareness of the long-term nature of the process: "you hear something from childhood many, many times and you keep it in your brain and you store it, then when you hear it later, it becomes easy to learn."

Participants seemed conscious of the difficulty caused by words irrelevant to their field. In support of the influence of area of study (item 2), Helin (19, Software Engineering) stated that "studying vocabulary of your interests will make it easier for you to learn … it is easier to learn words if they are about your department."

Word factors

The second research question was about the contribution of word factors to test-takers' perception of difficulty in the tests. An analysis of the items and comments relating to the

Table 2. Survey results for the test-taker factors (N=588).

Item no.	Item	Mean	Std. deviation
1	Attending the Preparatory School has made it easier for me to learn academic vocabulary.	2.50	1.42
2	If the academic word directly relates to my own field of study, it is easier to learn.	3.77	1.17
3	It is easier for me to learn words I need for academic purposes if I encounter the words outside the university context.	4.30	1.08
4	It is easier for me to learn academic words if they are used in L1.	4.13	1.20
14	The amount of time allowed to learn academic words makes it easier for me to learn.	3.45	1.18

form and meaning of the vocabulary is presented in Table 3, which shows the mean scores and standard deviations for these survey items.

Frequency of exposure (item 5) scored highest, indicating that test-takers perceived words encountered infrequently to be the most difficult to learn. The second strongest response was to item 12, indicating that students tended to find words with multiple meanings more difficult. In contrast to expectations, abstractness (item 7) was found to have the lowest mean score, 2.51; among 588 participants, 108 strongly disagreed, 214 disagreed, 147 were undecided, 92 agreed, and only 27 strongly agreed. The relationship between multiple meanings and abstractness emerges as an important theme in the discussion.

Only three word factor items obtained mean ratings of greater than three; frequency of exposure (item 5) ranked the highest, whereas the only highly ranked form/meaning items were multiple meanings (item 12), and pronunciation (item 8). Less influential factors were those relating to form: number of syllables (item 6), grammatical form (item 9), and affixes (item10); and to meaning: abstractness of meaning (items 7), abstractness of context (item 11), and formality of context (item 13). Perceived difficulty was overwhelmingly associated with frequency of encounter, rather than the formal aspect of words, with the exception of pronunciation.

A possible explanation for the lack of influence of abstractness (item 7) was given by Nurşah (22, Psychology). She considered that abstract words are familiar through content studies: “I am studying psychology, so does that mean I will not be able to learn the academic words of my department? No, it does not!” She also emphasizes that concepts in her own field are more abstract than, for example, those in Computer Engineering. Gizem (21, International Relations) highlighted a possible relation between abstractness and multiple meanings: “when (a word) is abstract, it generally covers more than one meaning, so it is hard.”

Contrary to expectations and the literature, most formal features of words seemed to present little difficulty. The following comment from Gizem (21, International Relations) is revealing:

Table 3. Survey results for the word factors (N=588).

Item no.	Item	Mean	Std. deviation
5	Academic words are harder to learn if not encountered often.	4.13	1.06
6	It is more difficult to learn academic words with multiple syllables.	2.72	1.23
7	It is more difficult to learn academic words that are more abstract.	2.51	1.09
8	It is harder to learn academic words which are more difficult to pronounce.	3.05	1.23
9	The word form (noun, adjective, verb, etc.) affects the difficulty of learning academic words.	2.56	1.19
10	If the word has prefixes or suffixes it is more difficult to learn.	2.84	1.18
11	The degree of abstractness of the context of the word makes it more difficult to learn academic words.	2.61	1.11
12	Academic words are more difficult to learn if they have multiple meanings.	3.29	1.18
13	The formality of the context in which the word is used makes it more difficult to learn.	2.86	1.14

> once you realize that this word is hard because it is long, you pay special attention to that word… and therefore you remember it easily. Because it is now a special word for you. The same with pronunciation. You tell yourself, "Oh, I cannot say this word" and you pay special attention, and you say it (Smiles).

This explanation suggests that difficulties that are related to sound or form can be overcome by various learner strategies, particularly by verbal repetition. Another insight was given by Murat (30, International Relations), who stated that 'if the word has prefixes and suffixes, it is easier to learn', contradicting expectations of the perceptions of difficulty regarding length.

Test factors

Although outside the scope of the survey, the role of test factors emerged during the interview, particularly in relation to the effect of item placement on test-taking motivation. Two participants mentioned the effects of difficult words encountered early in the test, with one of these, Helin (19, Software Engineering), stating:

> The exam finishes the way it starts. If the first part of the vocabulary exam is hard, or if you believe that it is hard, no matter how easy the other parts are, you get a lower grade. But, on the other hand, if the first part starts with words that you already know, and if you can do that part, somehow, all the channels in my brain are opened and I am able to do all. So, yes, how you perceive the test is important.

Merve (19, Public Relations) saw difficulty in terms of the presentation of the tested word, implying the role of collocation, stating that "the ease or difficulty of vocabulary tests depends solely on the context that the unknown word is being used. When you look at the sentence, if there is an unknown word … then it is harder to guess. But, if you understand the sentence and know the meaning of all the words, finding the answer is easier."

Although difficulties were reported in regard to specific parts of the test and the contextualizing sentences, one participant, Gizem (21, International Relations) had a solution: "First, I do the questions that I know in each part of the test, and then if there are a few left, I try to do them, and most of the time I decide on the choice which sounds the best. It is a kind of an instinct actually."

Discussion

The results suggest EAP vocabulary testing is subject to different factors from general language testing, and that item writing at this level requires particular skills and knowledge. Milton (2009, p. 43) emphasizes that by understanding where learning is concentrated, test writers can target the words test-takers are likely to know, resulting in a good estimate of knowledge, rather than selecting items at random from different frequency bands (in this case, of the AWL). To obtain this knowledge of test-takers, it is important for test writers to be able to 'envision the skills and competencies of students' through training and discussion (van der Watering & van der Rijt, 2006). In the light of the data, and the need to understand the test-takers' knowledge, the following section draws four implications for EAP vocabulary item writers, respectively related to the social and educational background of test-takers, the academic context, the features of words, and the tests themselves.

Implication one: The social and educational background

Item writers should be aware that the test-takers' social environment and educational background has a major influence on perceptions of difficulty. In relation to social background, Corson (1997, p. 682) argues for the key importance of exposure to the 'culture of literacy' outside the institution, which, he argues, has been seriously neglected in vocabulary research. Corson (1997) argues that experiences outside and before formal education are key to academic success. Thus, the lack of influence of the Preparatory School in the current study contrasts with the perceived importance of secondary school English language education, highlighted by one participant, in line with Corson's (1997, p. 694) observation that "a long childhood history of L2 learning … can strengthen several of the factors that affect the learning of academic English." Table 1 shows that the majority report more than five years' learning experience in addition to attendance at the Preparatory School, suggesting that both are required. EAP vocabulary item writers should therefore not assume that intensive preparatory courses alone are sufficient for familiarity with the items of the AWL.

Unfortunately, it would be almost impossible for item writers to gain detailed insight into the social context and past educational experience of each individual test-taker, which would vary greatly according to personal circumstances. Nevertheless, one aspect of the social context is very relevant to item writing. A link is implied between the two

highest scoring item 3 (social context) and item 4 (cognates) because the entry of English words into L1 is likely to be important in encountering words in the non-academic context. De Groot and Keijzer (2000) provide evidence that cognates are easier than non-cognates. It will be essential for item writers to have an understanding of the extent to which L1 incorporates words from L2, and how this can affect perceived difficulty. However, awareness is also needed of the extent to which test-takers are able to recognize different cognates.

Implication two: The academic context

Item writers should understand the major role of the academic context (i.e. field of study) on perceptions of difficulty. Unlike social factors, insights into factors relating to field of study are much more accessible to items writers through content knowledge. It is important for item writers to be aware of the fact that academic words may have highly specific uses within particular disciplines, which, in some cases, may be regarded as technical rather than sub-technical vocabulary (Hyland & Tse, 2007, p. 249). Item writers who specialize in teaching EAP in particular fields are likely to have some insight into the specialist meanings within those fields. However, ideally, test item writers would have a subject-specific background, or work in collaboration with, or at least with input from, subject specialists.

Item writers would also benefit from an understanding of the role of subject-specific reading in acquiring academic vocabulary. The relatively high score for item 2 highlights the perceived importance of familiarity with words through repeated contextualized encounters in academic texts. Incidental learning is of key importance in learning academic vocabulary: each encounter leads to very small increments in learning; however, this cumulative process strengthens mental representations over time (Wesche & Paribakht, 2009). While incidental learning is a slow process, the knowledge gained is suitable for word recognition in MCQs (Waring & Nation, 2004), such as those used in this context. Time also emerged as a factor representing a moderate degree of difficulty in the survey (item 14), further underlining the slowness of the process. Perceived difficulty will relate to the number of times a word has been encountered before a test, and at least 10 encounters in context are needed for basic knowledge (Webb, 2007). However, the frequency of encounter of AWL words may vary between fields; although proposed as suitable for all fields, the AWL has been criticized for bias towards Economics and Law (Hyland & Tse, 2007). In the current study, this is a possible source of perceived difficulty for participants in faculties other than Business Administration. Item writers should be aware that, at this level, the role of incidental learning will have a much greater influence on difficulty perceptions than classroom vocabulary instruction, owing to the lack of time available for the latter. However, they should also understand that the amount of exposure to the AWL through this process may vary according to field of study.

Implication three: The features of words

Item writers need an understanding of the perceived difficulty of factors that are related to words themselves, particularly in relation to the academic and social context. In this

study, perception of difficulty was not generally associated with the formal attributes. While this may be partly owing to participants' ability to use various strategies, including knowledge of the morphological structure of Greco-Latin vocabulary (Corson, 1997), interviewees reported that the area of study dominates perceptions of the difficulty associated with word factors, particularly with abstractness. In previous studies, abstractness has been associated with difficulty because of the relative infrequency of abstract words, and their lack of imageability (e.g. De Groot & Keijzer, 2000). However, in the current context, the perceived difficulty of abstractness appeared to be counteracted by the effect of the Context Availability Theory (Connell & Lynott, 2012), which states that contexts for more abstract words can be more easily created because of the learners' academic and social environment, as one interviewee emphasized. Nevertheless, as pointed out by another participant, abstractness may be an indirect cause of perceived difficulty; abstract words tend to have a wider range of meanings (Wang Tzu & Nation, 2004). Thus, the challenge may lie not in the abstract nature of the known sense, but in the potential for abstract words to have diverse senses, some not commonly used or known outside a particular field. Item writers should be aware of these key differences between general purpose and academic vocabulary, and, in particular, avoid assuming that abstractness itself necessarily results in higher levels of perceived difficulty.

Only two word factors were found to be influential: multiple meanings and pronunciation. Participants clearly perceived that difficulty was caused by polysemous words, and their different meanings according to field of study, which has been discussed in the previous implication. The second word factor found to cause perceived difficulty was pronunciation. Item writers may underestimate this factor in written tests where there is no need for oral production. Regardless of the purpose of learning, pronunciation is essential in enabling vocabulary items to pass into the long term memory (Baddeley et al., 1998). Murphy (2004) points to word stress as a neglected area in academic vocabulary teaching. Even in MCQs, item writers need to take into account the influence of irregularities in word stress, and form/sound correspondences. Moreover, phonological familiarity is the major contribution to the ease of learning cognates (De Groot & Keijzer, 2000), further underlining the importance of pronunciation. Like all word factors, the perceived difficulty caused by pronunciation at this level is closely tied to the social context, as pronunciation may be best learnt through contact with L2 speakers (Hansen Edwards, 2008).

Without an understanding of the relationship between words and the wider context, the teacher and test writers may focus excessively on features such as word length, form or abstractness in isolation. In other words, they may have "too much knowledge" (van der Watering & van der Rijt, 2006), in terms of the technical aspects. Such technical expertise can hinder a realistic view of test-taker difficulty perceptions because such aspects in isolation contribute only minimally to test-taker perceptions.

Implication four: The test

It is important to understand the relationship between perceptions of test difficulty and word difficulty. Test factors were a secondary focus of the overall study, which mainly explored perceptions of the tested knowledge. However, item writers aiming to balance

the difficulty of items over the whole test also need an understanding of how test factors influence perceptions of difficulty. Three points emerged: first, that perceptions of difficulty of different parts of tests can affect motivation and thus performance; second, that the contextualizing sentence itself can be a cause of perceived difficulty in MCQs; and finally, test-taking strategies play a role in reducing perceived difficulty.

In regard to parts of the test, according to Iwashita, McNamara, and Elder (2001), the effect of the order of items on perceived difficulty relates both to the test and the test-takers, since the perception of the whole exam can be influenced by its parts. As one interviewee pointed out, the perception of excessive difficulty at the beginning may cause demotivation, and thus, reduced language resources over the whole exam (Robinson, 2001, p. 32).

Second, an interviewee considered that perceived difficulty in the input (i.e. the sentence in which the word is presented) was more important than in the response (i.e. the word tested). However, this opinion was not reflected in the answers to survey item 11, which concerns the abstractness (and therefore, the perceived difficulty) of the context, as opposed to the word itself. This may be owing to the view that abstractness in general is not a cause of perceived difficulty if the subject area is familiar.

Finally, another interviewee described using test strategies to overcome the above-mentioned test effects. Strategy use has been reported in choosing options in MCQs; for example, Cohen (2006, p. 319) noted a number of approaches, including drawing on knowledge of the world, previous experience of tests, and making a best guess.

While test factors can influence perceptions of difficulty, both between and within test item types, it can be argued that the word itself is fundamental. For MCQ tests, variation in perceived difficulty between parts of a test is largely a result of the differences in the perceived difficulty of the specific words tested. The perceived difficulty of the contextualizing sentence is directly related to the collocation and context of tested words, and test strategies will only be necessary where test content is perceived as difficult. In other words, test item writers' understanding of the perceived difficulty of the words themselves is of key importance in creating balanced tests which avoid concentrations of items that are perceived as difficult in certain sections or inappropriate contextualizing sentences.

Conclusion

This research investigated test-taker perceptions of difficulty, revealing that these perceptions may be rather different from item writers' perceptions, especially if the latter are narrowly focused on technical aspects of decontextualized words. The study reveals that the EAP vocabulary-learning process, and hence, difficulty perception, is characterized by complex interactions of factors, dominated by the social and academic context. The study also has implications for understanding the role of pronunciation in difficulty perceptions.

The questionnaire developed for this study could be extended to understand further the perceptions of test-takers, especially in regard to the social context, which was found to have the most influence. Survey items could focus on different aspects of the social context and examine the respective influence of peers, family, L1, and media consumption on difficulty perceptions. Similarly, in future studies, it would be interesting to make the social context the focus of test-taker interviews, in line with Corson's (1997) calls for the investigation of

the social context in vocabulary learning. Another possible direction is to compare actual item difficulty with perceptions of difficulty, as identified by the questionnaire developed in this study. Such an approach has the potential to shed light on test performance, and also reveal how far test-taker perceptions of difficulty are reflected in their performance.

In regard to the academic context, this research highlights the importance of defining academic vocabulary for specific disciplines, particularly regarding those polysemous words whose various meanings are each associated with a particular field, in order to guide test writers and test-takers. It would also be interesting to replicate the study in other contexts to determine the extent to which the difficulty perceptions in this context are reflected in other cultural contexts. Finally, all these conclusions have implications for test item writer training, which should aim to develop judgments based on a realistic view of test-takers' knowledge in general, and their perceptions of difficulty in particular.

Acknowledgements

The authors would like to thank the current and former editor, and anonymous reviewers for valuable insights.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Anderson, N. J., Bachman, L. F, Perkins, K., & Cohen, A. (1991). An exploratory study of construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, *8*(1), 41–66.

Bachman, L. F. (2003). *Fundamental considerations in language testing* (6th ed.). Oxford: Oxford University Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998).The phonological loop as a learning device. *Psychological Review*, *105*(1), 158–173.

Baker, T. L. (1994). *Doing social research* (2nd ed.). New York: McGraw-Hill.

Bonaccio, S., & Reeve, C. L. (2010).The nature and relative importance of students' perceptions of the sources of test anxiety. *Learning and individual differences*, *20*(6), 617–625.

Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing*, *7*(1), 13–30.

Brown, A. (1993). The role of test-taker feedback in the test development process: Test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, *10*(3), 277–301.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319.

Classon, D. L., & Dormody, T. J. (1999). Analyzing data measured by individual Likert-Type items. *Journal of Agricultural Education*, *35*(4), 31–35.

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, *3*(4), 307–331.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465.

Corson, D. (1997). The learning and use of academic English words. *Language Learning*, *47*(4), 671–718.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign language learning and forgetting. *Language Learning*, *50*(1),1–56.

Dogancay-Aktuna, S., & Kiziltepe, Z. (2005). English in Turkey. *World Englishes*, *24*(2), 253–265. Retrieved from http://staff.neu.edu.tr/~cise.cavusoglu/status%20of%20English%20in%20Turkey.pdf.

Elliott, M. (2013).Test-taker characteristics. In A. Geranpayeh & L. Taylor (Eds.), Examining listening: Research and practice in assessing second language listening. *Studies in Language Testing*, 35 (pp. 36–76). Cambridge: Cambridge University Press.

Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of Foreign Language vocabulary learning. *Language Learning*, *43*(4), 559–617.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London: Routledge.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge Falmer.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference, 11.0 update* (4th ed.). Boston, MA: Allyn and Bacon.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity, and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281.

Hansen Edwards, J. G. (2008). Social factors and variation in production in L2 phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.). *Phonology and second language acquisition* (pp. 251–282). Amsterdam: John Benjamins.

Harrell, M. C., & Bradley, M. A. (2009). *Data collection methods: Semi-structured interviews and focus groups*. Santa Monica, CA: National Defense Research Institute.

Huhta, A., Kalaja, P., & Pitkanen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, *23*(3), 326–350.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long. (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell.

Hyland, K., & Tse, P. (2007). Is there an "Academic Vocabulary?" *TESOL Quarterly*, *41*(2), 235–253.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning*, *21*(3), 401–436.

Kim, H., & Elder, C. (2014). Interrogating the construct of aviation English: Feedback from test-takers in Korea. *Language Testing*, *32*(2), 129–149.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436.

Liu, J. Y., Chang, Y. J., Yang, F. Y., & Sun, Y. C. (2011). Is what I need what I want? Reconceptualising college students' needs in English courses for general and specific/academic purposes. *Journal of English for Academic Purposes*, *10*(4), 271–228.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*, 379–413.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual matters.

Murphy, J. M. (2004). Attending to word-stress while learning new vocabulary. *English for Specific Purposes Journal*, *23*(1), 67–83.

O'Sullivan, B. (2000).Towards a model of performance in oral language testing. Unpublished PhD thesis. University of Reading, UK.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London: Continuum International.

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

Robinson, P. (2001). Task complexity, task difficulty, and task production. *Applied Linguistics*, *22*(1), 27–57.

Scarcella, R. C., & Zimmerman, C. B. (2005). Cognates, cognition, and writing: An investigation of the use of cognates by university second language learners. In A. E. Tyler, M. Takada, Y. Kim & D. Marinova (Eds.), *Language in use: Cognitive and discourse perspectives on language and language learning* (pp.123–136). Washington, DC: Georgetown University Press.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, *36*(2), 145–173.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, *18*(4), 373–391.

Van de Watering, G., & Van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, *1*(2), 133–147.

Wang Tzu, K., & Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics*, *25*(3), 291–314.

Waring, R., & Nation, P. (2004). Reading and incidental vocabulary learning. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Writing and vocabulary in foreign language acquisition* (pp. 97–110). Copenhagen: Museum Tusculanum Press.

Webb, R. (2007). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Applied Linguistics*, *28*(1), 46–65.

Wesche, M. B., & Paribakht, T. S. (2009). *Lexical inferencing in a first and second language: Cross linguistic dimensions*. Bristol, UK: Multilingual Matters.

Xioa, Y., & Carless, D. R. (2013). Illustrating students' perceptions of English language assessment: Voices from China. *RELC Journal*, *44*(3), 319–340.

Appendix A: A sample list of AWL

Faculty of Engineering ENG101.

Listening Unit 1	Listening Unit 2	Listening Unit 3	Listening Unit 4
Component	Process (v)	Occurrence	Assign
Accuracy	Compile	Assemble	Bias
Locate	Interpret	Conceptualize	Evident
Detect	Transform	Incorporate	Obtain
Range	Formula	Approach (n)	Conduct (v)
Simulate	Purchase	Inference	Primary
Structure (n)	Convert	Rationally	Validity
Flexibility	Linkage	Integration	Currency
Conform	Deduction	Subsequent	Indicate
Uniform	Correspond	Ambiguity	Cite
Equip		Enable	
		Induce	
		Capability	
Speaking Unit 1	**Speaking Unit 2**	**Speaking Unit 3**	**Speaking Unit 4**
Automate	Assume	Principal	Reveal
Survey	Benefit	Incident	Trend
Portion	Assess	Aspect	Facilitate
Maintain	Expansion	Intensify	Abstract
Analyze	Issues	Advocate	Derive
Specify	Conclude	Enforce	
Significant	Involve	Proportion	
Index	Impact	Diversify	
Rational	Emerge	Displace	
Establish	Occupy		
Creator	Available		
Range	Ultimate		
Occur			
Discriminate			
Respond			

Appendix B: The questionnaire

Studying and learning vocabulary for English for academic purposes

Part A of the questionnaire below (Items 1–14) lists the factors that affect your perception of difficulty of the vocabulary sections of the ENG101 and ENG102 exams. Please read each item on the questionnaire very carefully and choose **only one** of the choices given on the scale.

Your efforts are very much appreciated and the information on this form will be kept confidential and used only for research purposes.

Thank you.

Name 1						
Name 2						
Personal Information						
1. Age:	A. 16	B. 17.	C. 18	D. 19	E. 20	F. Other
2. Gender:	A. Female B. Male					
3. Did you attend the Preparatory School?	A. Yes B. No					
4. How long you have been learning English?	A. 0–1 year B. 2–3 years C. 4–5 years D. more than 5 years					
5. Faculty:	A. Faculty of Economics and Administrative Sciences B. Faculty of Fine Arts and Design C. Faculty of Engineering and Computer Sciences D. Faculty of Arts and Sciences E. Faculty of Communication F. School of Applied Management Sciences, Culinary Arts and Management					

Survey: Factors affecting vocabulary test answer selection

Please state the extent that you agree or disagree that each factor below affects your ability to select the correct meaning of academic words in exams.

		Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1.	Attending the Preparatory School has made it easier for me to learn academic vocabulary.	A	B	C	D	E
2.	If the academic word relates directly to my own field, it is easier to learn.	A	B	C	D	E
3.	It is easier for me to learn the words I need for academic purposes if I encounter the words outside the university context.	A	B	C	D	E
4.	It is easier for me to learn academic words if they are used in L1.	A	B	C	D	E
5.	Academic words are harder to learn if they are not encountered often.	A	B	C	D	E
6.	It is more difficult to learn academic words with multiple syllables.	A	B	C	D	E
7.	It is more difficult to learn academic words that are more abstract.	A	B	C	D	E

		Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
8.	It is harder to learn academic words that are more difficult to pronounce.	A	B	C	D	E
9.	The word form (noun, adjective, verb etc.) affects the difficulty of learning academic words.	A	B	C	D	E
10.	If the word has prefixes or suffixes it is more difficult to learn.	A	B	C	D	E
11.	The degree of abstractness of the context of the word makes it more difficult to learn academic words.	A	B	C	D	E
12.	Academic words are more difficult to learn if they have multiple meanings.	A	B	C	D	E
13.	The formality of the context in which the word is used makes it more difficult to learn.	A	B	C	D	E
14.	The amount of time allowed to learn academic words makes it easier for me to learn.	A	B	C	D	E

This is the end of the survey. Thank you for your help.

Appendix C

Focus group interview questions:

1. Does age have an important effect on the difficulty of EAP tests?
2. Does the length of time spent studying English have an effect on the difficulty of EAP vocabulary tests?
3. Is your faculty an important influence on how difficult you perceive EAP vocabulary tests?
4. Is gender an important influence on how difficult you perceive EAP vocabulary tests?
5. Does attending the Preparatory School have an effect on how difficult you perceive EAP vocabulary tests?
6. Does GPA [grade point average] have an effect on how difficult you perceive EAP vocabulary tests?
7. How important are word factors such as abstractness, length of words, and pronunciation?
8. Are there any other factors, such as recency, that affect your perception of difficulty in EAP tests?
9. Did questions about academic vocabulary difficulty increase your awareness of the process of answering exam tasks?
10. How do you think your perception of an exam as easy or difficult affects your performance?