**“Teacher, is the exam difficult?”**

**Students’ perceptions of difficulty in vocabulary assessment tasks**

Nesrin Oruç Ertürk a,

Simon E. Mumford b,

**Abstract**

It is common practise to assess vocabulary through multiple choice and matching tasks. While these tasks may appear to be a straightforward test of vocabulary, many factors contribute to students perceptions of difficulty, which in turn may influence performance. This study aims to identify factors which influence student perceptions of difficulty through a questionnaire which measures students’ perceptions of difficulty related to the test taker factors. The study, set in an EAP context, deals specifically with academic vocabulary, which presents challenges in terms of word form, multiple meanings, and morphology. The results reveal that neither of age, length of studying English, or faculty had any effect on guessing the meaning and the perception of difficulty in vocabulary tests; however, gender is a factor at the perception of difficulty.

**Key Words:** vocabulary tests, test-taker perception, exam difficulty

**1. Introduction**

Multiple matching and multiple choice tasks are commonly used in second language vocabulary testing. In particular, multiple choice exams have been described as an efficient way of providing useful information about students’ ability in receptive skills (Brown & Hudson, 1998). These tests have the benefit of being objective, since there is only one correct answer, and therefore no judgement is required by the marker (Read, 2000, p. 76). However, in spite of the popularity of such tests, especially for vocabulary, Read (2000, p. 78) notes that there is little on-going research, although one aspect that has had a small amount of attention is that of difficulty, which is the main focus of the current research. The research on perceptions of difficulty for listening and speaking tasks seem to be more developed than that of vocabulary.

a Corresponding Author: School of Foreign Languages,Izmir Universityof Economics, Sakarya Caddesi No: 156, Balçova, 35330, Izmir, Turkey. Email: [nesrin.oruc@ieu.edu.tr](mailto:nesrin.oruc@ieu.edu.tr). Tel.:+90 232 4888174

b School of Foreign Languages,Izmir Universityof Economics, Sakarya Caddesi No: 156, Balçova, 35330, Izmir, Turkey. Email: [simon.mumford@ieu.edu.tr](mailto:simon.mumford@ieu.edu.tr) Tel.: +90 232 4888176

While multiple choice tests are relatively convenient and efficient ways to measure vocabulary, such test formats clearly involve more than simple vocabulary knowledge. As the aim of test writers is to produce tests that are as uniform as possible in terms of difficulty, it is important to understand difficulty from the students’ point of view. Difficulty is not the same as complexity, as it is directly related to the test-takers themselves (Robinson 2001), and has been described as “the mental ease or difficulty with which linguistic items are learned, processed or verbalised in the process of language acquisition and use” (Hulstijn & De Graaff 1994, in Bulte & Housen 2012, p. 23).

According to Bachman (2003, p. 13) the distinction between test characteristics, or facets, (such as time allowed, task types) on the one hand, and test taker factors (e.g. age, sex) on the other is important because the former are more open to manipulation by test writers, while the latter are fixed. Therefore, it seems not only to be important to understand the students’ perceptions related to these two different sources of potential difficulty, but also, to determine which of facets of tests that are open to control are perceived as being more influential. Thus, a distinction is drawn in this research between factors that influence the perceived difficulty of learning the words themselves, which relate mainly to linguistic aspects, but also to the educational context, and the factors that influence the process of selecting an answer in a multiple choice vocabulary exam (factors that relate to the test-takers).

The test-takers themselves are the focal point of the study because the design and modification of any test should take into account their perceptions. As Green and Andrade (2010) note, writers should take students views into account when revising test format, stating that “Not only do we need to understand teachers’ and students’ beliefs about teaching, learning, and testing, but also need to provide opportunities for them and other participants to ‘buy in’ to the reasons for the change. Any type of reform should require stakeholder participation and should focus on stakeholder learning as part of the reform.” (p. 325). This research therefore aims to focus on student perceptions of difficulty, both in order to understand their beliefs, and also to provide a learning opportunity for the stakeholders involved (i.e. students and test-writers). Understanding the perception of difficulty is potentially useful to teachers, test writers, and the students themselves. It has implications for learner training, since an understanding of the test situation has the potential to reduce perceptions of difficulty, and possibly help students improve their performance. For test writers, insights into perceptions of difficulty may help to improve the quality of tests, as they will be able to see items from the test takers point of view, and not just their own. An understanding of perceptions of difficulty may be useful to help teachers provide strategies and advice to students, which may cause them to reassess their view of the concept of difficulty, with possible benefits for the test taking process.

**2. Literature Review**

The literature review looks at test-taker factors, from various aspects. Below is a compiled collection of the selected studies from the literature.

* 1. *Test taker factors*

This research draws on a number of areas relating to test factors and perceptions of difficulty. Bachman (2003) discusses a number of test taker factors, including age, sex, native language, social and educational background. In the Korean context, Pae (2011) investigated gender difference in relation to EFL test performance, and found an interaction between item type and gender, finding males were more likely to do better on blank filling vocabulary tests, and females in reading comprehension.

In relation to variables affecting test-takers, Robinson (2001, p. 31) notes that task complexity is fixed, but task difficulty varies between individuals, and depends on the resource pool of each. Between learner variance is affected by difference in “available cognitive resources”. He notes two types of learner variables, firstly, affective variables, such as ‘confidence, motivation and anxiety’. These can change in the short term e.g. from task to task. The other types are ability variables, such as intelligence, aptitude, and cognitive style, which are more fixed. Both these types have an effect on the attention pool of students (p. 32). Robinson also notes that affective variables are difficult to measure because they can change very quickly, whereas ability variables are more fixed and thus easier to measure. In the current study, the effect of ability was measured through a questionnaire item relating to GPA.

This research also takes into account social factors affecting student performance, such as social and educational background. In particular, as this study is set in a second language medium university in an EFL context, another aspect of this research focuses on whether attending the university preparatory programme affects the perception of difficulty.

As well as increasing the understanding of student perceptions of difficulty for researchers, teachers and test writers, this study may also have the benefit of increased student awareness of factors that affect their performance in tests. McNamara (2001, p. 345) points out the importance of making students aware of their performance as an object, and “develop an emerging metalanguage about their performance.” An indirect outcome of this research is therefore likely to be increased student reflection by students on their performance in vocabulary test tasks, through the administration of a questionnaire, and for a smaller number of participants, the follow-up interviews.

This study therefore aims to analyse students’ perceptions of difficulty with regards to the factors that affect test-takers’ perceptions of vocabulary test tasks. This has a number of potential benefits for all stake-holders, including test-takers, test-writers, teachers and course coordinators, since insights gained can help students to reconceptualise difficulty, while giving test takers and teachers insight into factors that are perceived as creating difficulty, which can influence the way tests are produced.

**3. Methodology**

*3.1. Research context and site*

The results discussed in this article have been deduced from the study which aimed identify factors which influence EAP freshman students’ perceptions of test difficulty. The study aimed to measure students’ perceptions of difficulty and the variables that might affect their perceptions such as age, gender, field of study, etc.

The study took place in 2013 at a private university in Turkey, which provides English-medium instruction and a one year prep education before students start studying in their faculties.

*3.1.1. Test writing context*

This research was carried out within the specific context of the development of tasks for the academic vocabulary section of the Midterm and Final exams for the courses ENG 101 and 102, worth 20 percent of the total exam grade. These tasks and their weightings are also regularly reviewed. Furthermore, in the interests of producing balanced tests across six faculties, it is important to understand how various factors contribute to perceived difficulty, and how these affect the test takers’ approach to the tasks. Therefore, this research is part of a process of ongoing test task review and efforts to understand the factors that influence the difficulty of academic vocabulary test tasks.

The question pursued in this research is as follows:

1. To what extent do test-taker factors have any effect on guessing the meaning of EAP vocabulary and the perception of difficulty?

*3.2. Participants*

The participants of the study were all first year undergraduates of the university. All students were of an upper-intermediate (B2) English level (as a result of the prep education) and were taking the compulsory 4-hour-a-week ENG102 course in the year of 2012-2013 spring semester. All were Turkish and their first language (L1) was Turkish.

The students were chosen on the basis of convenience. In total there were 1392 freshman students taking this course. In order to collect data from 50 percent of the whole group, it was decided to distribute the survey to 100 students from each faculty. However, since there are in total 55 students of Culinary Arts and Management, all students of this department were included. In total, 588 students participated. The students range in age from 16 to above 20. Subjects were assured of their anonymity and right not to participate. No incentives for participation were offered. Below are details about the participants.

Table 1 Participants

Age 18 16

19 90

20 236

Other 246 n=588

Gender Male 345

Female 243 n=588

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prep Education Yes 538

No 50 n=588

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Length of Studying English 0-1 39

2-3 148

4-5 47

+5 354 n=588

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Faculty[[1]](#footnote-1) BA 104

FA 112

FE 102

FAS 112

FC 110

Cul. 48 n=588

*3.3. The Instrument*

The items on the instrument used for data collection was developed using information from the literature and the researchers’ own expertise. There were 46 items on the scale, the first 6 of which involved demographic information such as age, gender, faculty, etc. The other 39 items were about the students perceptions on vocabulary learning. The questionnaire had two subscales: Part A: Opinions about guessing the meaning of vocabulary. There were items like “the frequency of meeting the word is important” and “the number of syllabi in the word is important”. The second subscale was Part B: Factors affecting the learning of academic vocabulary which actually sought for the factors affecting the perception of difficulty, sample items included “It is harder to learn abstract academic words” and “It is harder to learn academic words which are harder to pronounce”. In Part A there were 22 items, followed by 17 items in Part B.

A multi-item (5-item) scale (Classon & Dormody, 1999)was considered most suitable for this age range. Therefore, the Likert-type response scale ranged from 1 (Strongly Disagree) to 5 (Strongly Agree) for each item. In order to test the reliability of the scale, the researchers conducted a pilot study, detailed below. The questionnaire was written in English and then translated into Turkish and conducted in Turkish, the participants’ mother tongue. The researcher’s translation was later checked by a faculty member in Translation and Interpretation Department.

*3.4. Focus Group Interviews*

Due to the large group of participants in the survey, it was not possible to include all the freshman students who answered the questionnaire to the interviews. Therefore, 10 students, at least one student from each faculty (5 female and 5 male) were chosen for the focus group interviews in a random yet strategic way to ensure representation from all six faculties.

Interviews were held in spring 2013 in July after the students had filled in the questionnaires. Interviews were conducted at the university where the study was organized. An interview guide was used and the interviews were semi-structured (Harrell & Bradley, 2009) built from existing theory and researchers experiences in the field. Only one of the researchers conducted the interviews for two reasons. First, so as not to increase the anxiety level of the students and second, the interviews, like the questionnaire, were held in students L1 to be able to collect as much as information as possible and only one of the researchers shared the same L1 with the students. It was stressed that anonymity would be maintained.

*3. 5. Procedure*

The students filled in the questionnaires during their class hours. The teachers (*n*=13) who distributed these questionnaires were invited for a meeting where they were given the necessary information both about the study and the survey. The purpose of the study was explained, and they were asked to react to the participants’ questions with, ‘Just choose the answer that you think is best – try to write down whatever comes to mind as comprehensively as you can’. The students were given 20 minutes which was decided as a result of the time students spent in the pilot study.

**4. Data analysis and Results**

The data was analyzed using SPSS .15. T-test and One-way ANOVA were computed and the results are presented below for each research question posed.

In the first part of the questionnaire the participants were asked to state some personal information such as; age, gender, length of studying English, their faculty and ENG101 GPA band. Furthermore, since the questionnaire had two subscales: A) opinions about guessing the meaning of vocabulary, B) factors affecting the learning of academic vocabulary, in order to see the effect of each variable listed above on these subscales, ANOVA was used to compare the mean scores and therefore to answer the first research question which asked:

RQ1: Do test taker factors have any effect on guessing the meaning of EAP vocabulary and the perception of difficulty?

It was seen that the participants’ age was not a significant factor for them to guess the meaning of academic vocabulary in an EAP context (F=1.85; df= 3; p>.136); however, age was a significant factor in their perception of difficulty (F=3.77; df= 3; p<.011); indicating that students at different ages perceive the difficulty of a test item differently.

Results show that neither of length of studying English, or faculty had any effect on guessing the meaning and the perception of difficulty in vocabulary tests, i.e. It does not matter how long they have studied English (F=.87; df= 4; p>.482) and at which faculty they study (F=1.34; df= 4; p>.252).

The other variables were gender, prep education and ENG101 GPA. The results reveal that gender is a factor at both guessing the meaning (F=6.82; df= 2; p<.001) and the perception of difficulty (F=8.22; df= 2; p<.000) in EAP tests. The results give no indication of which gender is perceived to be more successful or to find vocabulary tests easier; however, we can, using results as evidence, claim that gender affects both the process of selecting the meaning of a word and the perception of difficulty of test items. Another variable for which a positive result was obtained was the prep education, among our participants there were students who were exempt from prep education before they study at their departments. Therefore we wanted to see if having a one-year prep education had any effect on guessing the meaning and the perception of difficulty. The results show that prep education has a significant effect on both. There is a difference between the students who had prep education and who did not. Having a prep education has a significance value of .001 which is statistically significant (F=6.82; df= 2; p<.001) for guessing the meaning of unknown words in an exam. The same was true for perceiving a test item as difficult (F=10.77; df= 2; p<.000). It is possible to say that having prep education has a significant effect on the perception of difficulty for the students who have participated to that study, presumably decreasing the perception of difficulty.

The last factor studied was the GPA of ENG101 course. This study was conducted in the second semester of the academic year and this course (English for Academic Purposes) is given in two different semesters, namely fall and spring. Fall semester course is coded as ENG101 and spring semester course is coded as ENG102. Interestingly, although the results reveal that ENG101 GPA does not have any positive or negative effect on guessing the meaning since it has a significance level of .457 (F=.952; df= 6; p>.457), the same variable has a significant effect on the perception of difficulty (F=7.84; df= 6; p<.000) which means students’ first semester grades affect the way the perceive a test item as difficult or not.

Table 2 ANOVA results for each variable

Variable Guessing the meaning sig. Perception of difficulty sig.

Age 18 19 20 Other 18 19 20 Other

Mean 2.92 3.18 3.17 3.23 2.71 2.88 2.98 3.06

.136 .011

Gender Male Female

Mean 3.26 3.09 .001 .000

Prep Education Yes No

Mean 3.22 2.93 .001 .000

Length of studying English[[2]](#footnote-2) 0-1 2-3 4-5 +5 0-1 2-3 4-5 +5

Mean 3.06 3.24 3.22 3.18 3.06 3.06 3.00 2.94 .482 .252

Faculty[[3]](#footnote-3) BA FA FE FAS FC Cul. BA FA FE FAS FC Cul.

Mean 3.09 3.20 3.20 3.22 3.22 3.25 3.01 2.94 2.95 2.98 3.00 3.16 .487 .366

ENG101 GPA 0-45 46-60 61-75 76-90 +90 0-45 46-60 61-75 76-90 +90

Mean 3.32 3.22 3.24 3.18 3.09 3.24 3.27 3.11 2.92 2.77 .001 .000

Other than the personal data which was asked at the beginning of the questionnaire, beneath each subscale there were items to test the participants’ opinion on the effect of factors which related to the test taker. As a part of the analysis, the items which tested the test taker factors as a subscale were also studied. Table 3 below shows a detailed analysis of the results. Among test taker factors, we included aspects such as age (Item #35), gender (Item #36), educational level (Item #37), prep education (Item #38) relatedness to individual’s field (Item #42), time allowed (Item #43), encountering the word outside the school context (Item #44) and L1 effect (Item #45). Among these eight factors, the participants seem to agree that the greatest influence was encountering the words outside the school context. Of 588 participants 340 strongly agreed it is easier for them to learn words they need for academic purposes if they encounter the words outside the school context (mean=4.30).

The least affecting factor among human factors was gender. 365 participants believed that gender was not a factor in identifying a test as difficult or not. Only 21 agreed and 22 totally agreed with the statement that gender makes it more difficult to learn academic vocabulary. Another generally chosen factor was “If the academic word directly relates to my own field of study, it is easier for me to learn.” Not surprisingly, the participants feel that academic words which are close to their own field of study are easier to learn (mean=3.77).

Other than those least chosen and mostly chosen factors, there are some other factors which were chosen moderately by the participants: age, educational level, preparatory school education, the amount of time allowed to learn academic words, and whether these academic words are used in Turkish or not.

Table 3 Test taker factors affecting students’ perception of difficulty \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Item # Item Mean Std. Deviation Sum

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

35 Age makes it more difficult to learn academic vocabulary. 2.02 1.19 1185

36 Gender makes it more difficult to learn academic vocabulary. 1.62 1.01 954

37 Education level makes it more difficult to learn academic vocabulary. 2.49 1.34 1467

38 I did not attend the Preparatory school, and this makes it more difficult

for me to learn academic vocabulary. 2.50 1.42 1455

42 If the academic word directly relates to my own field of study, it is easier

for me to learn. 3.77 1.17 2201

43 The amount of time allowed to learn academic words makes it easier

for me to learn. 3.45 1.18 2009

44 It is easier for me to learn words I need for academic purposes if

I encounter the words outside the school context. 4.30 1.08 2516

45 It is easier for me to learn academic words if they are used in Turkish. 4.13 1.20 2412

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

N=588

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**5. Discussion**

*5.1. Test taker factors*

*5.1.1. Area of study*

The students show awareness that different departments have different vocabulary needs. Helin (19, Software Engineering) states 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.’

At first it seems as if the test takers are confusing EAP, the general Academic Vocabulary taught on these courses, with ESP, the technical vocabulary needed for the subject. However, it is becoming increasing recognised that not everything in the AWL will be equally useful for students from all departments. Each branch has specific needs, and work has been done on the specific needs of Engineering students (Ward 1999), and Agriculture students (Martinez, Beck & Panza, 2009). The problem seems to be that the AWL is a general list for all subjects, but increasingly, students are studying EAP in subject specific contexts, and that in this case, only the high frequency AWL words are likely to be of value to all groups, with the less high frequency words more suitable only for specific subjects (Satarsyah, Nation & Kennedy 1994). It is also becoming recognised by course designers that not only content but also language academic language including academic vocabulary should be selected according to students’ study needs (Yurekli, 2012) and this seems to be recognised as a potential source of difficulty by the test takers themselves.

*5.1.2. Age*

Experienced related factors seemed to dominate the task takers factors, specifically age, previous learning experience, and the involvement in the university preparatory school. There seems to be conflicting evidence concerning the role of age. In spite of its lack of importance in the questionnaire, it emerged as a factor in the focus group. However, rather than being related to the development of cognitive skills, which has been the main focus of research, age appeared as a motivational factor.

According to Helin (19, Software Engineering) ‘…and your concerns about your future. How is it gonna be? What am I gonna be? When you realize that you have to be something then you start to give importance… I actually passed this course but I am taking it again as an Improve to get a higher grade and therefore to increase my GPA.’

A similar approach is taken by Murat (30, International Relations), who is considerably older than the others in the group. He states ‘as you get older, you give more importance to English. I, myself, for example, now know that I should learn more vocabulary and therefore do my best to learn more.’

This suggests that age has an important effect on motivation, and that age brings the need to be successful student to ensure a good career. Therefore, generally speaking the older the student, the more effort they are likely to put into learning, and the more importance they attach to exam success.

*5.1.3. Previous experience*

Perhaps related to age is the previous experience of learning English. Helin (19, Software Engineering) states ‘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.’ This statement reflects recent research, which suggests that repeatedly seeing words in context is necessary for a word to become known. However, this participant may have overestimated the time needed to learn a word, with Webb (2007) suggesting that meeting a word 10 times is enough for “sizeable gains in knowledge.” In contrast, the participant sees the process as taking years, and suggests a rather passive approach to learning in which long exposure is necessary before a word can be learned properly. This seems to represent a rather behaviorist view of language learning, in which long term habits equal learning, whereas in fact, a more active role of the learner will make learning much less time consuming. Such a view may also produce an exaggerated view of difficulty.

Nurşah states that the university prep school helps in terms of general English, *but not at all* *for academic English.* This seems to suggest that academic and General English are seen as two different types of vocabulary. In fact it has been suggested by Billuroglu and Neufeld (2005, p. 7) that the Academic word list has been ‘institutionalised’ as ‘advanced’ vocabulary, particularly in prep schools in universities in non-English speaking countries, whereas in fact, they argue, it is the way the words are used (collocation and context) which make them academic. To associate these words as advanced vocabulary may give a false impression that the words themselves are inherently different and difficult.

*5.1.4. Gender*

The questionnaire pointed to this not being a factor, and it was not expected to be important. However one test taker mentioned a specific instance:

It is a very important factor for my department. I am studying at Software Engineering and I am a girl but most of my friends are boys and as you know boys are very much interested in computer games. So these boys have been playing computer games since they are young and when they see a word or phrase they say “Yes, I know this from X game”. But I am a girl and I have never played computer games in my life so I do not know these words. So, yes, in that sense gender is important in my department.

This anecdotal evidence may have some truth in it. Palmberg (1988) has described the benefits of computer games on language learning, particularly vocabulary acquisition. This extract shows how different factors, in this case, Gender and area of study can combine in unexpected ways, highlighting the complexity relating to the differences between individual test takers, and therefore the factors influencing the test taking process.

**6. Limitations of the Study**

There are a number of limitations to this study. First, in order to increase validity some other data collection methods could have been used. In future studies, the same research questions can be answered using different qualitative and quantitative data collection tools (comparison of students’ grades with the results of questionnaires, reflective reports, etc.). This rich blend of data collection methods might present some other valuable data to help us investigate the phenomena in hand.

Second, studies conducted in varied local contexts (different cities, different universities –private/state) can be more meaningful if they are backed up with similar studies in different contexts. Perception of difficulty is person specific; therefore, collecting more data from more participants in different contexts can only increase reliability and present more reliable results.

**7. Conclusion**

This research has revealed the complexity of the concept of perceived difficulty of academic vocabulary test tasks. Not only are there a wide range of test, word and test-taker factors involved, but these interact in unpredictable ways. For example, gender may play a greater role in different areas of study. Also, aspects of word profile like length of word, abstractness, and pronunciation interact with test-taker factors in ways that mean they do not always increase difficulty, and may actually make it easier for words to be retained.

On a more global level, test-takers seem to have a strong awareness that some academic words are more useful, and more familiar to them than others, and that these words will be easier for them than others that are not directly related to their department. This accords with recent research into vocabulary needed for specific disciplines, and relates to the way words can have different meanings in different fields.

The members of the focus group were able to articulate a sophisticated view of the factors involved in perceptions of task difficulty, involving the order of tasks and the relative difficulty of each. They also showed that the difficulty of the context of the word (ie the sentence) was as important as the word itself. Regardless of whether test-writers are attempting to increase or decrease the level of difficulty, it is essential that they take into account students perceptions of difficulty, which may be different from their own. It terms of implications for classroom practice, this study suggests that it is important to demystify the AWL, and highlight the fact that these are not inherently difficult or advanced words.

**References**

Bachman, L.F. (2003). *Fundamental Considerations in Language Testing.* Oxford University Press : Oxford.

Brown, J. D., & Hudson, T. (1998). Alternatives in language testing. *TESOL Quarterly* *32*(4), 653-675.

Billuroglu, A. & Neufeld, S. (2005). In search of the critical lexical mass: How 'general' is the GSL? How 'academic' is the AWL? Retrieved 30-8-13from www.lextutor.ca

Bulte, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A Housen, F. Kuiken & I.Vedder. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. John Benjamins B.V Amsterdam

Green, B. A. & Andrade, M. S. (2010). Guiding principles for language assessment reform: A model for collaboration. *Journal for English for Academic Purposes*  9, 322-334.

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

Hulstijn, J.H. & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge?: A research proposal. *AILA Review* *11*, 97-112.

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

Martinez, I. A., Beck, S. C., & Panza C. B. (2009). Academic vocabulary in agriculture research articles: A corpus based study. *English for Specific Purposes* *28*(3),183-198.

McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing 18*, 333-349.

Pae, T-I. (2012). Causes of gender DIF on an EFL language test: A multiple-data analysis over nine years. *Language Testing 29*, 533-554.

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.

Satarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal* *25*(2), 34-50.

Ward, J. (1999). How large a vocabulary do EAP Engineering students need? Reading in a foreign language. Retrieved from <http://eric.ed.gov/?id=EJ601601> 27-8-13

Webb, S. (2007). The effect of repetition on vocabulary knowledge. *Applied Linguistics* *28*(1), 46-65. Doi: 10.1093/applin/am1048

Yurekli, A. (2012). An Analysis of curriculum renewal in EAP Context. *International Journal of Instruction* *5*(1), 49-68.

1. BA= Faculty of Business Administration

   FA= Faculty of Fine Arts and Design

   FE= Faculty of Engineering and Computer

   FAS= Faculty of Arts and Sciences

   FC= Faculty of Communication

   Cul= Culinary Arts and Management [↑](#footnote-ref-1)
2. In terms of year [↑](#footnote-ref-2)
3. BA= Faculty of Business Administration

   FA= Faculty of Fine Arts and Design

   FE= Faculty of Engineering and Computer

   FAS= Faculty of Arts and Sciences

   FC= Faculty of Communication

   Cul= Culinary Arts and Management [↑](#footnote-ref-3)