

# Administering and Evaluating an English proficiency test for study abroad candidates

Bradley Perks

## **Introduction**

In the era of globalization and the emergence of English as the Lingua Franca, the number of students studying abroad in English speaking countries is increasing dramatically. Admission requirements for the study abroad applicants in these English speaking countries require sufficient scores in various English proficiency tests. There is a wide range of English proficiency tests available, for example TOEIC, TOEFL, IELTS, and Cambridge ESOL, all of which measure the learners' mastery of the traditional four language skills: listening, reading, writing and speaking. There is a need for a proficiency test to measure non-native university student's ability to use and understand English at the university level. Furthermore, there is need to develop certain academic sub skills such as the ability to listen to lecturer deliver a speech in English, to take notes, paraphrase and summarise content, plus express their opinion is needed.

## **Purpose**

The aim of proficiency test is evaluate student's mastery of English speaking university. More specifically it designed to measure the learners' mastery of the traditional four language skills: listening, reading, writing and speaking. The first component will measure the participants'abilily to listen to an academic lecture in English and conversational dialogue delivered in English from native speakers at a native pace. The second component will measure the participants'abilily to read long English passages whilst applying reading techniques such as skimming and scanning to understand specific information and make inferences to understand the author's viewpoint in a timed situation. The third component will measure the participants ability to write a descriptive and short academic essay on academic content. The fourth component will measure the participant' s ability to competently participate in a dialogue with a native English speaker.

Focusing on specific English proficiency skills such as

- Reading
- Listening
- Speaking
- Writing

Focusing on specific academic sub skills such as

- Taking Notes
- Paraphrasing
- Summarizing

- Participate in class discussions (tutorials) about content

### **Content of the test**

The students are from an international university in Japan and will undergo either a semester or yearlong study abroad program in a foreign English-medium university. This factor was considered in the content of the test, in particular the listening and reading sections content were taken from topics on arts & humanities, practical and social science. Authentic sources were used for the reading section. For example Section 3a used a passage from *A Rough Guide to South Africa* travel guide and Section 3b used a passage taken from *Our Revolution: A Future to Believe In* novel. According to Lewkowicz (2000) these sources are authentic because they are intended for the general public consumption.

Whereas academic reading material was used in Section 3c, passages were taken from an Australian first year Sociological textbook. This was selected to evaluate whether candidates could understand academic reading passages actually used in Australian university. The listening sections 2a, 2b and 2c were recorded prescribing to authentic notion. The audio was authentic because they contained natural features of speech such as varying speeds of delivery, pauses and fillers (Blau 1990). The words per minute varied within each sections and common English language fillers were used such as ‘errr, ummm, let me see’ to recreate an authentic listening exercise.

## **The Pilot Study**

This a pilot study, which means it will be assessed for its feasibility, and if feasible it will be part of a larger study. As Van Teijlingen and Hundley (2002) state pilot studies are a crucial element of a good study design. Conducting a pilot study does not guarantee a final study or success in the main study, but it does increase the likelihood. The intended purpose of this study is to test the research instruments effectiveness on measuring the four language skills: listening, reading, writing and speaking. Developing and testing the adequacy of research instruments in the pilot study will hopefully act as a troubleshooting process to ultimately provide more valid findings in the main study. Which in this case it will provide internal validity of the questionnaires. In order to fulfil the internal validity, the pilot study and main study' s questionnaire will be administered in exactly the same way.

The test subjects completed the proficiency test in the form a self completion test each subject was given 4 hours to complete it, to recreate an authentic proficiency test condition. Their test scores were recorded and given to them, for their record. The feedback on the test itself was recorded on the program Google Docs, using the edit function the subjects could comment on the follow criteria;

Ambiguous and difficult questions

Questions that are too long

The completed tests were checked to see if all questions were answered also the time to complete each section was recorded during the test to determine whether certain sections or questions need re

wording, shortening or omitted. This process was conducted to discard all unnecessary, difficult or ambiguous questions Peat et al. (2002)

Pilot studies may also try to identify potential practical problems in following the research procedure. For example, one of the participants was unfamiliar with the research instrument, the program Google Docs and was unwilling to download and use this program, so an alternative paper version was administered. Other problems such as internet access and incomplete tests can also be identified and precautionary procedures or safety nets can be devised.

### **Test specifications**

Both the test candidates N (2) and the control subject N(1) commented on the deemed excessive length of the reading section 3b and section 3c which contained 457 words 795 words respectfully. A likely conclusion to draw from this response is that the candidates perhaps lack the academic reading sub skills to skim and scan large reading sections in order to read quickly and efficiently. Despite this response, the length of the reading section was not edited because as Hughes (2003) states proficiency tests are designed to test 'expeditious' reading skills where test candidates are forced to answer questions in a timed setting. However, to alleviate participants reading fatigue, paragraphs could be numbered, to facilitate skimming and scanning the text more quickly. Plus, reading section question could explicitly indicate which paragraph the corresponding answers are in to increase participant response rate.

## **Instructions**

This pilot study revealed a possible source of ambiguity in the wording of the multiple choice section, the instructions state to chose the ‘correct’ answer, however replacing this wording with ‘best’ is a less ambiguous choice. An argument could be made that there could be more than one technically correct answer in a multiple choice question, so make it clear and explicit the wording was changed to the ‘best’ answer. Ambiguous instructions can according to Hughes (2003) be a source of inaccuracy, furthermore this can mislead and present problems for comprehension which can in turn compromise the test’s reliability. Ensuring clear and explicit instructions in test designs can eliminate these problems according to Hughes (2003).

## **Representativeness of the sample reliability and validity**

It was evident to see that the initial test design lacked a sufficient representativeness of the sample, meaning it needed more variety in the source material and testing items. As Lee and Greene (2007) state that by providing more tasks in a test design this produces a greater representation of the candidates’ ability and equates to test validity. As a result, the revised test included a greater diversity of texts and testing tasks. The initial test mostly limited to vocabulary items in the grammar/vocabulary section, so a conversational cloze section (Section 1b) written in grammatically correct natural language.

Adjustments were also made to the listening section to achieve representation of the sample. A longer listening task was recorded in the

revised test, sections 2a and 2c featured two Australian university students talking about an academic topic and a 9-minute lecture on an academic topic. This task was specifically compiled to test the candidates' ability to understand spoken English in a formal lecture setting and also in a casual conversational setting. A university student is required to attend lectures, so the ability to listen and comprehend a spoken passage longer than 10 minutes is necessary according to Brindley and Slatyer (2002). The listening section achieves an accurate representation of the sample by testing actual language skills necessary for students wishing to attend English medium universities.

The reading section was also altered to achieve greater diversity. The content of section 3a was taken from a travel guide and 3b was taken from an English novel, which added to the variety of different genres and styles. Furthermore, the integrated essay which required participants to use a variety of language skills such read a text, listen to an audio excerpt and then write an essay based on that input was removed due to validity reasons. As Lee and Greene (pg. 78, 2007) state "the skills of listening and reading should not impact on the candidates writing ability". Writing sections 4a, 4b and 4c were added as a replacement. Section 4a was replaced by a descriptive writing task, section 4b was replaced with a written job application and section 4c was replaced with an apology letter. The objective of making these alterations was to fulfill greater detail in the specification of content, according to Lissitz and Samuelsen (2007) the greater the detail in specification of content the more valid the test is likely to be. Furthermore, enacting a suitable representation of the sample also constitutes 'content validity, which Lissitz and Samuelsen (2007) explain' as a valid specification of the skills or structures.

## **Scoring procedures**

The tests mostly administered closed ended multiple choice questions, which Pimsleur (1968) states is objective, plus “rapid, reliable and economical” according to Bachman (1990). Also open ended answers were also administered in the test, the questions prescribed to Hughes (2003) statement to allow the answers to be unique. As a result, in reading section 3c the questions starts off as a closed answered question where there can only be one choice i.e. “What does it refer to?”. Then develops into an opened ended question i.e. ‘What are the stages of cognitive development?’. These questions were separate, however related in the following questions; ensuring that a correct response on one item depends on correctly answering another item (Hughes 2003).

Whereas, scoring of the speaking and writing sections was less rapid and economical due to the open ended nature of writing section, for example section 4a and 4b require a written 150 word summary and sections requires a 75 word responses. An important factor when marking this test is not to be subjective, but to be objective as possible. Adhering to the objective writing and speaking aims and scoring scale, ensures that objectivity is remained. As Brown and Bailey (1984) state providing consistent measures of precisely the abilities needed achieves valid scoring.

## **Conclusion**

The English proficiency effectively measured the student’s abilities to understand likely course material at an English medium university. Academic topics on natural and social sciences and social science plus

arts & humanities were included. Authentic material was used, such as newspapers articles, dialogue based on real conversations, novels and real university textbooks. The proficiency test evaluated the participants ability to read a university level textbook reading passage, write an essay that conforms to academic writing conventions, produce a monologue and participate in a focused discussion. These criteria can provide the recipients of the test with information about the candidates' ability to successfully complete a study program. Piloting this proficiency test for study abroad participants allows further research to be conducted without ambiguous, unnecessary and inconsistent questions. To allow a more valid main study on English proficiency tests the research instruments effectiveness on measuring the four language skills.

### References

- Bachman, L. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press, 18-53.
- Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL quarterly*, 24(4), 746-753.
- Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning*, 34(4), 21-38.
- Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19(4), 369-394.
- Hughes, A. (2003) *Testing for Language Teachers*, 2nd edn. Cambridge: Cambridge University Press.
- Lee, Y. J., & Greene, J. (2007). The Predictive Validity of an ESL Placement Test A Mixed Methods Approach. *Journal of Mixed Methods Research*, 1(4), 366-389.
- Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language testing*, 17(1), 43-64.
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational researcher*, 36(8), 437-448.

- Peat, J., Mellis, C., & Williams, K. (Eds.). (2002). Health science research: a handbook of quantitative methods. Sage.
- Pimsleur, P. (1968). Language aptitude testing. In Language testing symposium: A linguistic approach (pp. 98-106).
- Van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, 16(40), 33-36.

## **Appendix 1 Test Specifications**

### **Content**

### **Operations**

#### Grammar/vocabulary

- Knowledge of grammar/vocabulary and set phrases
- Lexical knowledge and general structure of language

#### Listening

- Understand spoken English in conversation and lecture settings
- listen for gist, detail, function, purpose, topic and opinion

#### Reading

- Understand specific information
- Make inferences to understand the authors viewpoint
- Skim and scan

#### Writing

- Listening to a lecture
- Take Note
- Quote author (s)
- Paraphrase
- Summarize
- Write essays in an English academic situation, this prescribes to tasks which are based on real-life use of the targeted ability (Bachman,

1990)

### Speaking

- Speak about personal experiences and opinions
- Prepare a monologue based on a given topic

### **Types of text**

#### Grammar/vocabulary

- C-Test - second half of the word is deleted. This allows a wide range of topics, styles, and levels of ability to be tested (Hughes 2003)
- Conversational cloze - Fill the gaps with no hints. Every 8th or 10th word is deleted to assess the candidate's ability to process lengthy passages of language (Hughes 2003).

#### Listening

- Multiple choice with 4 choices, 1 possible answer with 3 distractors (Hughes 2003)

#### Reading

- Understand academic reading passages similar to college textbooks

#### Writing

- Integrated essay, featuring a reading passage followed by an audio lecture on the same topic. This integrates several parts of language elements, measuring reading and writing, together with understanding and listening (Hughes 2003).

#### Speaking

- Interview
- Discussion

### **Addressees of texts**

Writing, Reading, Listening

- University students aged between 18-30

Speaking

- Researcher

**Lengths of texts**

Writing

- Integrated essay of 150-225 words

Writing, Reading, Listening

- Single paragraph texts or dialogues, up to 650 words

**Topics**

Typically encountered at the university level

Speaking

- Dialogue based on personal topics (neighbors, shopping, movies, pets et cetera)
- Monologue on a conversation topic (favorite artist, prized possession, best holiday et cetera)

Readability

- Within a senior high school/ advanced intermediate level

**Structures**

**Dialect, accent, style**

Standard international English, American English is also acceptable from candidates

**Speed of processing**

Read at 300 words per minute

Listening to native English speakers, delivering at normal speed

Write at 40 words per minute

**Structural range**

Unlimited

**Vocabulary range**

- General academic and some technical words (definition supplied)

**Structure, timing, medium and techniques**

Test structure: Five sections

Grammar / vocabulary

Section 1a - Lexical knowledge and understanding of the text to select the best word or phrase

Section 1b - Knowledge of the structure of the language using articles, auxiliaries, prepositions, pronouns, verb tenses

Listening

Section 2a - Conversation on campus between two students - listening skills for purpose, detail, inference and pragmatics

Section 2b - Lecture on either arts & humanities or practical or social science - listening for main idea, organization, details and point (argument)

Reading

Section 3 - Reading on either arts & humanities or practical or social science - reading for main idea, purpose, cause, details and author's opinion

Writing

Section 4 - Read a short passage then listen to a short lecture and take notes from both write a short essay

Speaking

Section 5a - Introduction and interview. General questions about familiar

topics such (pets, sports, movies, et cetra)

Section 5b - Individual long turn. Monologue on a particular topic with content focused prompts and questions from the examiner.

**Number of items**

Section 1 - 20

Section 1a - 10

Section 2 - 9

Section 2a - 4

Section 2b - 5

Section 3 - 8

Section 4 - 1

Section 5 - 2

Section 5a - 1

Section 5b - 1

Number of items 40

Timing

Section 1a - 10 minutes

Section 1b - 10 minutes

Section 2a - 10 minutes

Section 2b - 15 minutes

Section 3 - 30 minutes

Section 4 - 25 minutes

Section 5a- 5 minutes

Section 5b - 5 minutes

TOTAL: 110 minutes

**Appendix 1 (The tasks that candidates have to be able to carry out)**

## Operational definition

### Listening

- Anticipate the purpose of the lecture/ tutorial  
(identifying key words from the speaker to figure out the purpose)  
(how this prepares you to receive information)
- Know what to ignore  
(key words/phrases commonly used to signal un important information)

### Taking notes

- How to abbreviate  
(common conventions / symbols)  
Explain the purpose of using abbreviations to document large amounts  
of content without having to write everything down

### Types of text

#### Writing Academic essays

- Grammar
- Paraphrase
- Quote author (s)

### Addressees of texts

Candidates are expected to be able to write and speak to native speakers of the same age and status

Produce writing texts prescribing to Australian university academic standards

### Length of texts

500 words

### 15-minute interview

Topics

Typically encountered at at the university level

Readability

Passages from academic texts and answer questions.

Dialect, accent, style

Australian English speakers speaking at a native speed in a formal education setting

Speed of processing