

Investigating the lexical validity of Pearson Test of English Academic

Introduction

A study was conducted by Kieran O’Loughlin of Melbourne University with the aim of investigating the lexical validity of Pearson Test of English Academic (PTE Academic). This study took place during the field testing of the test items, in the course of which up to 300 test taker responses were collected for each item. This research summary provides an advanced overview of the main results he found.

An important quality of any test of academic English is what may be termed lexical validity: the extent to which the vocabulary occurring in and elicited by the test, is representative of the vocabulary that test takers will encounter and be expected to produce, in real-world academic contexts.

The nature of academic vocabulary has been widely researched, notably by Paul Nation and co-workers (Xue and Nation, 1984; Nation, 1990; Sutarsyah, 1993; Sutarsyah, Nation and Kennedy, 1994), whose work resulted in, amongst other achievements, the compilation of the Academic Word List (AWL)(Coxhead 2000). The AWL is based on a 3.5 million word corpus of academic texts. It lists those words which occur frequently in texts from a range of academic disciplines (so excluding technical terms that are exclusive to one field of study).

Research methodology

For his study O’Loughlin singled out three item types, all designed to assess writing ability. These are classified in Figure 1 according to the nature of the prompt, this is the written or spoken text to which the test taker has to respond, and of the response required.

Item type	Prompt	Response
Summarize written text	Written text	One-sentence written summary
Summarize spoken text	Spoken text	Paragraph length written summary (50–70 words)
Write essay	Essay topic	Essay (200–300 words)

Figure 1: Three PTE Academic item types assessing writing skills

The item prompts and the test taker responses were analyzed in order to answer the following questions:

- What is the proportion of academic vocabulary found in the item prompts?
- What is the proportion of academic vocabulary found in test taker responses?
- What is the relationship, with respect to academic vocabulary, between the item prompts and test taker responses? In other words, if an item prompt contains a high proportion of academic vocabulary does it typically elicit a response that also contains a high proportion of academic vocabulary?
- What is the relationship between the proportion of academic vocabulary in essay responses and item scores? In other words, if a test taker writes an essay that contains a high proportion of academic vocabulary, is it reflected in the score?

Findings

Figure 2 shows the average percentage of AWL words found in prompts and test taker responses for each item type, as well as the correlations (Pearson r) between these proportions. The results show that broadly similar percentages of academic vocabulary were found in both prompts and responses, with the highest proportion of academic vocabulary being found in the first item type (summarize written text). According to Paul Nation, authentic academic English texts typically contain at least 4% of AWL words. The results show that, according to this criterion, the test is academic in quality, both in respect of the language that it presents to test takers, and the language which it elicits.

Item type	Mean % AWL tokens: prompt	Mean % AWL tokens: responses	Correlation prompt/response
Summarize written text	7.24	7.59	0.892
Summary spoken text	5.93	6.37	0.679
Write essay	5.86	5.37	0.726

Figure 2: Percentage of Academic Word List tokens in item prompts and responses

What is particularly evident, however, is the close relationship between prompts and responses, with correlations between 0.679 and 0.892. The correlation is lowest for the second item type (summarize spoken text). This is perhaps not surprising, as in this item type the prompt is a recording of spoken text, giving test takers no opportunity to directly copy written words into their responses. The highest correlation is found for the first item type. This also is not surprising as an accurate one-sentence summary will likely use words from the input text. The high correlation between prompt and response for this item type is shown graphically in Figure 3.

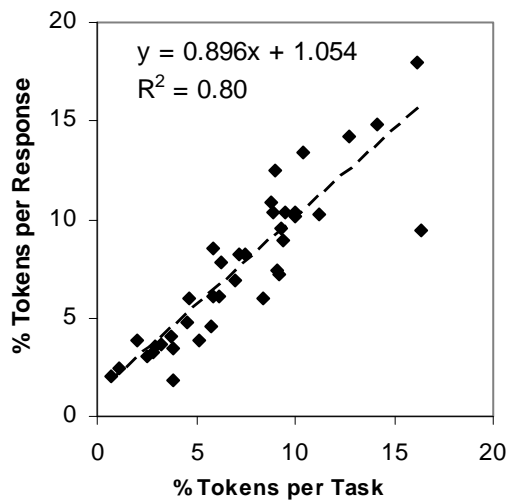


Figure 3: Summarize written text items: Percentage of AWL words in prompt and responses, item by item.

Here it can be seen that although there is considerable variation in the percentages of academic vocabulary found in the item prompts (from little more than zero to more than 16%), the percentages of academic vocabulary in the corresponding responses mirrors this closely, showing that a prompt that is academic in character does elicit responses that are also academic in character.

With regard to item scores, a range of aspect or trait scores are used in PTE Academic. For this study, however, the essay responses were assessed using a test-construct independent global rating, i.e. ratings based on the scale for Overall Written Production from the CEF (Council of Europe, 2001, 61–62). Test takers’ essays were assessed by expert markers in relation to the upper four levels of this scale. Figure 4 shows mean AWL-measurements for responses grouped by CEF level.

CEF level	Mean % AWL tokens	Mean % AWL types
Below B1	3.76	9.32
B1	4.76	8.59
B2	5.72	10.22
C1	6.42	15.21
C2	7.93	17.08

Figure 4: Percentage of AWL words measurements for responses by CEF level

The results reveal a strong relationship between the proportion of academic vocabulary and the quality of test taker responses in terms of an independent global rating. In other words, test takers whose writing is academic in character, at least with respect to vocabulary, perform better on a global scale of functional writing ability.

Conclusion

The findings of this study indicate that, with respect to these item types, the test is academic in character. They also show that there is a strong relationship between the percentage of academic vocabulary in item prompts and in the quality of the language these prompts typically elicit.

References

- Council of Europe (2001) Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.
- Coxhead, A. (2000) A New Academic Word List, *TESOL Quarterly*, 34(2), 213–238.
- Nation, I.S.P. (1990) Teaching and Learning Vocabulary. New York: Newbury House.
- Sutarsyah, C. (1993) The Vocabulary of Economics and Academic English. Unpublished MA thesis, Victoria University of Wellington, New Zealand.
- Sutarsyah, C., I.S.P. Nation and G. Kennedy (1994) How useful is EAP vocabulary for ESP? A corpus-based case study. *RELC Journal* 25, 2: 34–50.
- Xue Guoyi and I.S.P. Nation (1984) A university word list. *Language Learning and Communication* 3: 215–229.

Further PTE Academic research initiatives

PTE Academic is a new international computer-based academic English language test, delivering real-life measures of test takers’ language ability to universities, higher education institutions, government departments and other organizations that require academic-level English. The test will accurately measure the Listening, Reading, Speaking and Writing ability of test takers who are non-native speakers of English and who want to study at institutions where English is the principal language of instruction. Launching in October 2009, PTE Academic is being developed and will be offered to students in collaboration with and with endorsement from the Graduate Management Admission Council® (GMAC®), owner of the Graduate Management Admission Test® (GMAT®).

To inform the development of PTE Academic, Pearson has assembled an international team of leading researchers and experts in language studies and applied linguistics to direct the following research initiatives.

Investigating the Academic Character of Written Production in PTE Academic

Description The findings of this research project support PTE Academic’s claim to be a test of academic proficiency by examining the extent of academic vocabulary test takers used in their written responses during the first field test. Researchers collected data from PTE Academic item types that required written responses, and analyzed them in relation to the Academic Word List compiled by Averil Coxhead, PhD, in 2000.

Sensitivity Review of PTE Academic Items

Description This project was set up to develop sensitivity review guidelines and offer recommendations for all items in the PTE Academic item bank. An international Bias-Sensitivity Review Panel evaluated which items in the PTE Academic item bank may be sensitive and advised on a statistical evaluation of item bias.

Rater Training Report

Description This project compared and evaluated the rater training programs for PTE Academic during the first field test at test centers in Australia, the United Kingdom, and the United States. Researchers delivered data-driven suggestions for improving future supervisor and rater training programs.

Concordance Study

Description Researchers will publish a concordance table comparing test takers’ scores on the TOEFL iBT and IELTS to their PTE Academic scores. This research will help universities compare admission criteria for non-native English speakers wishing to enter tertiary education.

Rater Training Materials

Description Experts in language test development and applied linguistics developed rater training modules and materials for PTE Academic.

Glossary

Traits a distinguishing or marked feature of an item
Item task in the test consisting of instructions, question, or prompt, answer opportunities, and scoring rules
Item type form of an item defined by the presentation, the nature of the task, the skills addressed, and the scoring rules
Test taker examinee, candidate

About Pearson

Pearson (NYSE:PSO), the global leader in education and education technology, reaches and engages today’s digital natives with effective and personalized learning, as well as dedicated professional development for their teachers. This commitment is demonstrated in the company’s investment in innovative print and digital education materials for pre-K through professional learning, student information systems, and learning management systems, teacher development, career certification programs, and testing and assessment products that set the standard for the industry. The company’s respected brands include Scott Foresman, Prentice Hall, Addison Wesley, Benjamin Cummings, the Stanford Achievement Test, the Wechsler family of assessments, SuccessNet, MyLabs, PowerSchool, SuccessMaker, and many others. Pearson’s comprehensive offerings help inform targeted instruction and intervention so that success is within reach of every student at every level of education. Pearson’s commitment to education for all is supported by the global charitable giving initiatives of the Pearson Foundation. Pearson’s other primary businesses include the Financial Times Group and the Penguin Group.

For more information, go to www.pearson.com