Title of Project:
How Valid are Domain Experts’ Judgments of Workplace Communication? Implications for Setting Standards on the Occupational English Test (OET) Writing Sub-Test

Researcher:
Simon Davidson
University of Melbourne
s.davidson3@student.unimelb.edu.au

Research Supervisors:
Dr. Ute Knoch
University of British Columbia
Assoc. Prof. Cathie Elder
University of Melbourne

Project Summary

As part of the requirement to attain professional registration and be employed in Australia, International Medical Graduates (IMGs) need to satisfy a number of prerequisites that include demonstrating satisfactory English language proficiency and competent clinical skills/knowledge. Some concern has been noted that the minimum standards on tests used to evaluate language proficiency may be inadequate for proving effective workplace readiness. To better comprehend the validity of these apprehensions, this study examined the use of domain expert judges in determining performance standards for a specific purpose language (LSP) test: the Occupational English Test (OET). The OET is used to evaluate the English language writing, speaking, reading and listening abilities of health professionals; in this research project they are medical doctors who received their qualifications outside of Australia in non-English speaking contexts. The objective of the OET test is to establish health professionals’ abilities to communicate in English with other health professionals and patients in specific occupational related contexts. The writing task on the OET test requires test-takers to write is a letter of referral that is based on a set of case notes that have been provided. The OET writing sub-test’s minimum standards were set via a procedure that elicited insights from suitable medical practitioners who had experience with relevant workplace communication demands and could determine levels of competency regarded as acceptable for this specific use. Setting suitable performance standards on such an assessment is of vital importance. If the standards are set too low, the well-being and safety of patients may be endangered by health professionals who aren’t able to communicate effectively. However, if the standards are set too high, communicatively competent test candidates could be deprived of certification and, consequently, not be allowed to work in regions where their skills/knowledge are greatly needed.

The purpose of the study was to ascertain the minimum levels of English language written proficiency that were deemed suitable by domain experts for effective performance in the workplace. Another consideration was to take into account the extent that health professionals, without linguistic training, are able to attend to issues of language and communication independently of clinical competence,
The International Research Foundation
for English Language Education

which is their area of expertise. What was the basis for the decisions made by subject-matter experts (i.e., medical doctors) and how strongly did these underlying reasons parallel the communication construct that the OET is intended to measure? A previous study (Manias & McNamara, 2016; Pill & McNamara, 2016) investigated these issues in regard to the OET speaking subtest, while the current study concentrated on the OET writing subtest, which is to date underexplored area. This doctoral research is an independent part of a larger validation study of the OET.

The study was designed to answer four research questions:

1. What criteria do standard-setting participants use as a basis for their decisions in judging writing responses and to what extent are these decisions language-based?
2. Is there any variability between judges in what they attend to while setting standards?
3. How do standard-setting judges view the process and outcome of the standard-setting procedure?
4. What occupational specific standards (cut scores) do doctors set on the Occupational English Test (OET) writing sub-test?

To explore answers to the research questions, 18 health professionals (GPs, specialists and medical educators) were recruited to participate in standard-setting workshops, which were intended to elicit judgements about particular performance level allocations for the task. In addition, reasons and features for why participants deemed a writing sample to be eligible for a passing grade were investigated. The study utilized a mixed approach to data collection, using both qualitative and quantitative methods. A pilot study was also performed to choose a suitable standard setting procedure from two potential methods. To obtain additional understanding of the foundation of the standards being established, a sample of participants completed verbal reports in the form of a think-aloud protocol (TAP) was elicited. The health professionals’ observations from the workshops and verbal reports were coded thematically and inter-coder reliability checks were carried out. The qualitative analysis revealed that the participants took into account seven main features of the writing samples in deciding on their judgments of performance levels: task fulfillment, content, organization, expression, presentation, professionalism, audience recognition, and a separate category of ‘other’ for aspects that could not be accommodated in the above themes. The study focused on not only the establishment of new standards for the OET writing sub-test but also on (1) the effectiveness and validity of the method itself, (2) the participants’ comprehension of and confidence in the procedure, (3) panellists’ consistency in the method’s used, and (4) the “indigenous assessment criteria” (i.e., the values that underlie everyday judgments of performance by domain experts in actual workplace situations (Jacoby & McNamara, 1999) that participants used in assessing the characteristics of the writing samples.

Prior to new passing standards and the “cut scores” being calculated, a many facet Rasch analysis using FACETS software (Linacre, 2017) was undertaken to consider any variation in subject-matter experts’ performance level judgments regarding their being overly severe, lenient, or inconsistent. The subsequent quantitative analysis generated a slightly more severe passing standard than the existing one. This result paralleled the findings of similar standard-setting studies on the OET using the Analytic Judgment method (AJM) (e.g., Knoch, Elder, Woodward-Kron, Flynn, Manias, McNamara, Zhang Ying & Huisman, 2017; Pill & McNamara, 2016). The new standards were contrasted with present OET cut scores, which showed a higher “fail” proportion than the existing data set. The more stringent passing standard founded by subject-matter experts in this study could be interpreted as supporting suggestions
and views that the current standard is not high enough and that some IMGs who are not yet communicatively proficient are, nevertheless, being employed in Australian work environments with inadequate written communication skills.

The qualitative analysis additionally examined whether subject-matter experts are capable of judging language ability independently from other professional skills/knowledge as is required by Australian federal government conditions. Some domain experts’ decisions (a minority on the whole) were swayed by assessments of test candidates’ clinical competency, which is outside the construct of communicative competence as circumscribed by the OET. Yet overall, the qualitative results indicated that domain experts were undeniably focusing on textual features associated with what the OET is meant to assess. The fundamental consideration of whether subject-matter experts, without language assessment training, are well placed for establishing the standards in a language for specific purposes (LSP) test such as the OET was taken into account. Despite the fact that some participants’ decisions departed to some degree from the present OET writing sub-test criteria, validity confirmation gathered in this study confirmed, for the most part, that the ensuing new standards, resulting from subject-matter expert involvement, were acceptable. The validity outcomes of this study’s findings for the OET writing sub-test, and for LSP testing more generally, were considered by means of a new argument-based validity framework created by Knoch and Macqueen (in preparation).

The most noteworthy features of this study are that first, a formal standard-setting procedure has been carried out with domain expert involvement on the OET writing sub-test, giving rise to a methodically developed range of cut scores for explaining test candidate performance. In addition, further evidence has been gathered concerning how well non-linguistically trained subject-matter experts are able to assess language performance and address facets of language and communication independently (or not) of their evaluations of professional competency. In this regard, the findings have offered new confirmation for the case that some language testers have proposed that language knowledge and content knowledge in LSP performances are inseparable. Furthermore, in keeping with contemporary approaches to test validation, the study has framed the standard-setting discussion in the context of a validity argument which has not been performed before with an LSP test. Another practical contribution of this study is that it has suggested more justifiable and defensible performance standards for the OET writing sub-test than those that are presently being used as the origin of the current OET standards has not been publicly available or been accessible to scrutiny for some time.
References


Hunt, K. (1965). *Grammatical structures written at three grade levels*. NCTE research report No. 3. Champaign, IL: NCTE.


