Test Development of an Email Response Task for Academic Writing Assessment

Randall S. Rebman
Northern Arizona University

Abstract

This research involves the prototyping of an email request writing task as part of an EAP writing placement test. This study aims to explore how a broader coverage of the construct of academic writing can be attained adding a pragmatic task to the current test. Very few studies have looked at developing pragmatic tasks for assessing academic writing. The development of this test task seeks to fill this gap. An additional benefit of developing this test task is the potential of ensuring more effective email communication between non-native speakers of English and university faculty through test washback. A complication that was highlighted in the prototyping of the new test task was the issue of task complexity. Task design, scale design and construct irrelevant factors were posited as some of moderating variables effecting undesirable outcomes of the piloted test task. Implications for developing future email writing tasks are also covered in this report.

Background

This paper reports on the test development of an email request writing task for use in a norm-referenced test of academic writing. Developing this writing task is part of an effort to expand the content coverage of the target language use (TLU) domain (Bachman & Palmer, 2010), which for this English for an Academic Purposes (EAP) program, involves different genres of writing used in university settings. The current academic writing test being used by the Program of Intensive English at Northern Arizona University includes two writing tasks that are used to make placement decisions into the university and the English language program. The two tasks currently being used for this test are an independent invention task and an interdependent text-based task from a reading source. However, arguments made by writing assessment researchers have articulated the need for also adding a third task for large-scale writing assessments such as the TOEFL iBT. According to these testing professionals, the third task should be an "interdependent situation-based task from listening or reading sources" Cumming et al, 2000, p. 10). The prototype email writing task is an example of such a task. Thus the development of this email writing task is part of a process to add more content coverage of academic writing tasks in the testing of the overall construct of academic writing.

Research Questions

1. Can an email request task be used to rank-order examinees in a meaningful way as part of a norm-referenced test?

2. Can two raters score the email request task consistently?

Methods

The participants included in the prototyping of the email task include 103 international students from the countries of Saudi Arabia, China, Kuwait, Japan, Korea, and Africa. The average age of participants ranges from 17-24 years old. The proficiency level of the participants ranged from beginner to advanced levels. Raters of the students' performance on the email task were teachers in the Program of Intensive English at Northern Arizona University. Three of the raters were non-native speakers of English, while the other three were native speakers of English. All of the raters had at least one year of ESL/EFL teaching experience.

The test task was developed with the intention of creating a scenario that emulates the target language use task characteristics as much as is practically possible in an assessment setting. The selection of a request for test task aligns with the type of speech act identified as frequently used in university settings in emails between students and faculty (Youn, 2009) and that has also been researched in non-assessment settings (Biesenbach-Lucas, 2007; Chen, 2006; Bloch, 2002). The task prompt requires that the examinee provide a self-introduction, identify a problem, and ask for advice from a professor. Time allotted for this task is fifteen minutes, which should allow for ample time for the examinee to construct a response to the best of their ability.

The email rubric operationalizes particular subconstructs. For example, researchers have defined register as consisting of three components: "the situational context, the linguistic features, and the functional relationships between the first two components" (Biber & Conrad, 2009, p. 6). For the assessment of register features on the email task, the bands 4 and 5 on the rubric integrate this interaction between context and

use in that appropriate language is used and a personal connection is established. Other subconstructs are also operationalized through the rubric bands, but they are not consistently measured throughout each band of the rubric. The reason for this is related to the approach taken in the construction of the rubric itself.

A score report form is given to students based upon the different descriptors in the rubric. When a student receives a certain score, the examiner checks the items on the checklist that correspond with the summed score between two raters who used the 6-point scale. Thus an examinee can receive a summed score ranging from 0-10. Raters undergo a benchmark training session on sample responses that I have scored along with one other rater who teaches at the Program of Intensive English. If raters are not within one point on the rubric with their given score for a response, then they discuss the differences and attempt to come to an agreement. In the case of disagreement a third rater is brought in to help reach an agreed upon score.

Results

The mean scores for the email task (6.90) show that examinees scored higher on this task than the integrated task (4.91) and independent task (4.66). These values reflect the summed scores of examinees on each task on a scale of 0-10. The higher scores on the email task show that the research question of weather the new email task could rank-order examinees appropriately was not achieved. Inter-rater reliability between raters on the email task was strong (0.92) using a Spearman-Brown correlation coefficient. Despite the strong correlation between raters, the undesirable dispersion of test scores means this test task is not desirable for including on a norm-referenced test.

The unfavorable results mean that the test task must be revised or the rubric revised and the task scored again. The results also highlight the issues of task complexity in creating a pragmatic task for assessment purposes. Creation of a more complex task is likely to better rank-order examinees in a way that would be desirable for making norm-referenced test. However, until future prototyping of the task is undertaken and variables on the scale and task are manipulated, it will be difficult to determine what exactly causes an email task to be increase in levels of complexity. An additional interpretation that might be made from the results of this test development project is that it might be a better task for criterion-referenced tests than norm-referenced tests.

References

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Biber, D., & Conrad, S. (2009). *Register, genre and style*. Cambridge, UK: Cambridge University Press

Biesenbach-Lucas, S. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning and Technology, 11*, 59-81.

Bloch, J. (2002). Student/teacher interaction via email: The social context of internet discourse. *Journal of Second Language Writing, 11*, 117-134.

Chen, C-F. E. (2006). The development of e-mail literacy: From writing to peers to writing to authority figures. *Language Learning & Technology, 10*(2), 35-55.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). TOEFL 2000 writing framework: A working paper (TOEFL Monograph No. 18). Princeton, NJ: Educational Testing Service.

Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press

Youn, S.J. (2009). *Developing a task-based assessment of EAP pragmatics*. Unpublished manuscript, Department of Second Language Studies, University of Hawai'i, Honolulu, HI. Retrieved from http://scholarspace.manoa.hawaii .edu/handle/10125/20180