

Investigating the Validity and Reliability
of an Abbreviated Placement Test and a Full Placement Test

Jacqueline Church
Northern Arizona University

Abstract

This study addressed the difference between a full version of an IEP placement test and an abbreviated version of an IEP placement test. All students took the full placement test at the beginning of the semester. The abbreviated version of the placement test was administered one week later, during the first week of instruction. This study was carried out using the Rasch model for analysis, using the program Facets (Linacre, 1996), with the facets being the examinees, items from each test, and raters for the speaking and writing domains. The data indicated that the test-takers did not perform similarly in each domain between the full placement test and the abbreviated placement test. Also, the items on the full placement test and the abbreviated placement test within each domain differed in terms of difficulty.

Keywords: IEP placement test, language testing, norm-referenced testing

Investigating the Validity and Reliability
of an Abbreviated Placement Test and a Full Placement Test

Background

While it is widely accepted that a sampling of ten items is required to adequately measure student ability on different item types (Miller, Linn, & Gronlund, 2009), this does not account for practical issues regarding the resources available to administer, score, and analyze test results. The PIE offers three semesters a year; fall, spring and summer. Fall and spring semesters have the largest enrollment and are 15-week semesters. Summer has a lower enrollment, and is a compressed 11-week semester. Typically, there is a low number of incoming international students for summer semester, as most arrive during the spring and fall semesters. Due to the low number of students who need to take the placement test during the summer, the PIE questioned the practicality of administering the full version of the placement test. Therefore, the PIE has developed an abbreviated version of the placement test. This abbreviated test is only used for placing students; students cannot use their scores on this test to matriculate. The abbreviated test includes four skills or domains, reading (26 items), listening (10 items), writing (1 task) and speaking (1 task). The test takes approximately 1.5 hours to administer, 2 hours to rate, 5 hours for data entry and analysis, and requires the participation of 4 to 5 staff members.

In terms of resources such as time, personnel, and money, an abbreviated placement test is appealing for a language program due to concerns of practicality. The purpose of the present study is to investigate whether the PIE abbreviated placement test is a valid measurement when compared to the full version.

Research Questions

1. Do examinees perform similarly in each domain (listening, reading, writing, and speaking) between the full placement test and the abbreviated placement test?
2. Are the items on the full placement test and the abbreviated placement test able similar in terms of difficulty?

Methods

The examinee data used for this study is from the full placement test administered in the fall semester of 2015. The data for the full placement test included 63 students; as there were three tasks for both the writing and speaking domains this resulted in 189 samples for both domains. The abbreviated placement test was administered to students one week later, during the first week of the semester and prior to instruction, thus, the student language proficiency should have remained static. Of the 63 test participants, 37 were male (59%) and 26 were female (41%). The first languages of the participants were Chinese (42, 67%), Arabic (19, 30%), Portuguese (1, 1.5%), and Spanish (1, 1.5%).

The instruments used for this study are the PIE full placement test and the PIE abbreviated placement test. For both tests, listening and reading items were dichotomous, and the speaking and writing tasks were polytomous. The scales for the polytomous tasks were the same score range and used holistic rubrics.

This study will be carried out using the Rasch model for analysis, using the program Facets (Linacre, 1996), with the facets being the examinees, items from each test, and raters for the speaking and writing domains.

Results

In answering the research question of whether examinees perform similarly in each domain between the full placement test and the abbreviated placement test, the data indicates that the test-takers do not perform similarly. The abbreviated test was easier for the listening test by over 0.50 logits, the reading abbreviated test was harder for the examinees by over 0.30 logits, the abbreviated speaking test was easier for the test-takers by 0.35 logits, and the abbreviated writing test was harder for students by over 0.60 logits. Based on these results, it is clear that the abbreviated placement test does not measure examinees consistently when compared with the full placement test.

Relevance to PIE and Second Language Learning

The purpose of this study was to address whether the abbreviated placement test measured students similarly as the full placement test. The research questions were focused on both the domains and the items. Based on these results, it is clear that the abbreviated placement test is not currently working appropriately for the PIE. However, certain domains were performing similarly across both tests, which indicates that a shortened test may be able to reliably predict student ability. In the future, the abbreviated placement test will be revised and then analyzed again.

References

- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.
- Barkaoui, K. (2013). Multifaceted Rasch analysis for test evaluation. In A. Kunnan (Ed.) *The companion to language assessment* (pp. 1301–1322). Cambridge: Wiley.
- Brown, J. D. (1989). Improving ESL placement using two perspectives. *TESOL Quarterly*, 23, 65-83.
- Coniam, D. (2006). Evaluating computer-based and paper-based versions of an English-language listening test, *ReCALL*, 18, 193–211.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Chicago, IL: Holt, Rinehart & Winston.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of c-tests. *Language Testing*, 23, 290–325.
- Fox, J. (2004). Test decisions over time: tracking validity. *Language Testing*, 21(4), 437-465.
- Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14(2), 113-138.
- Jamieson, J., Wang, L., & Church, J. (2013). In-house or commercial speaking tests: Evaluating strengths for EAP placement. *Journal of English for Academic Purposes*, 12, 288-298.
- Linacre, J.M. (1996). *Facets, version no. 3.0*. Chicago: MESA.
- Miller, M. D., Linn, R., & Gronlund, N. (2009). *Measurement and evaluation in teaching* (10th ed.). Upper Saddle River, NJ: Merrill, Prentice Hall.
- Purpura, J.E., Brown, J. D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. *Language Learning*, 65, 37-75.