

Parallel-Forms Reliability of the GEPT Speaking Test

Jessica R W Wu and Rachel Y F Wu

The Language Training & Testing Center

jw@lrtc.ntu.edu.tw; rwu@lrtc.ntu.edu.tw

Abstract

Examination boards, for security reasons, have to provide multiple versions of a speaking examination for each administration as well as across administrations. However, they rarely provide evidence of comparability across forms. The study aims to investigate the extent to which four forms of the General English Proficiency Test High-Intermediate Speaking Test (GEPT-HI-S) are parallel in terms of two types of validity: parallel-forms reliability and content validity. Performance data taken from the 2012 GEPT-HI-S administration were analyzed. In addition to more conventional statistical procedures, Multi-Faceted Rasch Measurement (MFRM) was employed to process test-takers' score data to account for the effect of variables associated with rater severity and test difficulty. Various checklists were also employed to compare the tests at task level in terms of content. Results of the study provide evidence of comparability across the four GEPT-HI-S forms. While this study presents only a partial validity argument for the GEPT-HI-S, it is hoped that this research will encourage other examination boards to apply methods appropriate to their contexts to gather validity evidence and demonstrate parallel-forms reliability.