ESTABLISHING THE VALIDITY OF
THE GENERAL ENGLISH PROFICIENCY TEST
READING COMPONENT THROUGH
A CRITICAL EVALUATION ON ALIGNMENT WITH
THE COMMON EUROPEAN FRAMEWORK OF REFERENCE

RACHEL YI-FEN WU

ABSTRACT

The present study aimed to establish an empirical framework for test validation and comparison of level-based test batteries and to identify parameters that are useful to explicitly describe different levels of reading proficiency examinations based on a critical evaluation of alignment of the examinations with the Common European Framework of Reference (CEFR). The scope of the study is limited to CEFR B1 and B2 levels. This study applied Weir’s (2005) socio-cognitive validation framework to examine various aspects of the validity of different levels of the GEPT in terms of contextual parameters, cognitive processing skills, and test results. The CEFR and two levels of a CEFR-aligned multilevel test battery, PET and FCE developed by Cambridge ESOL, served as external referents for a review of the similarities and differences between GEPT reading tests targeting CEFR B1 and B2 levels. To establish ‘situational and interactional authenticities’ (Bachman and Palmer, 1996), this study not only applied automated tools and expert judgment to examine ‘the degree of correspondence of the characteristics of a given language test task to the features of a TLU [target language use] task’ (ibid., 23), but also carried out what O’Sullivan (2006: 183) called ‘an a posteriori empirical exploration of test performance’ to gather evidence of interactional authenticity. The findings support the construct validity of the GEPT in general, but show that its cognitive validity needs to be enhanced by incorporating tasks which test expeditious reading operations. As regards the CEFR-alignment, the findings also show that procedures the Manual (CoE, 2009) recommends for linking an examination to CEFR levels do not produce sufficient evidence to demonstrate equivalence between different examinations that target particular CEFR levels. The results indicate that the GEPT Intermediate level
and PET, both of which target the B1 level, are equivalent, while the GEPT High-Intermediate level and FCE, which target the B2 level, are much different not only in terms of test results but also contextual features and cognitive processing operations.