World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:18, No:06, 2024

The Appropriate Number of Test Items That a Classroom-Based Reading Assessment Should Include: A Generalizability Analysis

Authors : Jui-Teng Liao

Abstract : The selected-response (SR) format has been commonly adopted to assess academic reading in both formal and informal testing (i.e., standardized assessment and classroom assessment) because of its strengths in content validity, construct validity, as well as scoring objectivity and efficiency. When developing a second language (L2) reading test, researchers indicate that the longer the test (e.g., more test items) is, the higher reliability and validity the test is likely to produce. However, previous studies have not provided specific guidelines regarding the optimal length of a test or the most suitable number of test items or reading passages. Additionally, reading tests often include different question types (e.g., factual, vocabulary, inferential) that require varying degrees of reading comprehension and cognitive processes. Therefore, it is important to investigate the impact of question types on the number of items in relation to the score reliability of L2 reading tests. Given the popularity of the SR question format and its impact on assessment results on teaching and learning, it is necessary to investigate the degree to which such a question format can reliably measure learners' L2 reading comprehension. The present study, therefore, adopted the generalizability (G) theory to investigate the score reliability of the SR format in L2 reading tests focusing on how many test items a reading test should include. Specifically, this study aimed to investigate the interaction between question types and the number of items, providing insights into the appropriate item count for different types of questions. G theory is a comprehensive statistical framework used for estimating the score reliability of tests and validating their results. Data were collected from 108 English as a second language student who completed an English reading test comprising factual, vocabulary, and inferential questions in the SR format. The computer program mGENOVA was utilized to analyze the data using multivariate designs (i.e., scenarios). Based on the results of G theory analyses, the findings indicated that the number of test items had a critical impact on the score reliability of an L2 reading test. Furthermore, the findings revealed that different types of reading questions required varying numbers of test items for reliable assessment of learners' L2 reading proficiency. Further implications for teaching practice and classroom-based assessments are discussed.

Keywords : second language reading assessment, validity and reliability, Generalizability theory, Academic reading, Question format

Conference Title : ICALPLLT 2024 : International Conference on Applied Linguistics, Principles of Language Learning and Teaching

Conference Location : Vienna, Austria

Conference Dates : June 20-21, 2024