

Provide a descriptive title

## Criterion-related Validity of the Wonder English Test Listening Section

Previous criterion-related validity studies of English for academic purposes (EAP) tests have often employed a summative measure such as grade point average (GPA) as the criterion of academic success. Despite the importance of such measures on various academic decisions made about students, the aggregated nature of such measures makes inferences about how nonnative speakers of English actually perform in the classroom difficult. Given that the interpretability of results of criterion-related validity studies hinges on the construct representativeness of the criterion measure (AERA/APA/NCME, 1999), it is imperative to seek an alternative measure of academic success that better reflects how nonnative speakers of English actually use the language in the target domain. As an attempt to address this issue, the present study investigated the criterion-related validity of the Wonder English Test Listening section by examining its relationship to a criterion measure designed to reflect language use tasks that nonnative speakers of English would often encounter in everyday academic life: academic lecture listening.

The design of the criterion measure was informed by 145 undergraduate and graduate students' survey responses on the frequency and the importance of various classroom tasks that require academic listening as well as the relationship between various class activities and assignments and the successful completion of academic courses. The criterion measure consisted of three video-taped lectures, each covering a different content area (physics, history, and psychology), and included tasks created by former professors of the relevant content area. These tasks reflected what the professors expected students to have comprehended during the lecture.

The criterion measure and the Wonder English Test Listening section were administered to nonnative speakers of English who were enrolled in undergraduate and graduate programs at eight institutions in the U.S. and Canada. Data from 221 participants were analyzed. Substantial correlations were observed between the criterion measure and the Wonder English Test Listening section score for the entire sample and for subgroups based on gender, academic level and field of study. The obtained Pearson correlation coefficients, ranging from .56 to .74, and disattenuated correlations, ranging from .62 to .82, were generally higher than validity coefficients typically reported in previous criterion-related validity studies of English language measures. Moreover, the mean scores on the criterion measure were compared across the High-, Intermediate- and Low-scoring groups (as reported in the Wonder English Test Examinee Score Report). The Intermediate-scoring group (i.e., those scoring between 14 and 21 on the Wonder English Test Listening section) represented the listening ability level required for admission to many undergraduate and graduate programs based on the results of previous Wonder English Test standard setting studies and cut scores reported by Wonder English Test score users. The results suggested that those participants who were classified in the High- or Intermediate-scoring groups scored, on average, nearly 50% or higher on the criterion measure, demonstrating reasonable comprehension of the academic lectures in the video sets. These results provided strong evidence supporting the criterion-related validity of the Listening section of this test.

Describe a rationale and significance of the study as part of the introduction

Provide sufficient details about the methodology, including participants, instruments, procedure and analysis, to help reviewers understand the study design

In the end explain key findings as well as conclusions and implications that might be drawn from the study findings