

Using Public Opinion to Inform the Validation of Test Scores

Tzur M. Karelitz

National Institute for Testing and Evaluation (NITE), Israel

Executive summary

Standardized tests, such as the Psychometric Entrance Exam (PET) are commonly the subject of a heated public debate. Public opinion can affect decision regarding the test's content and structure, its administration and scoring, and the interpretation and usage of test scores. Regardless of whether a test is psychometrically valid or not, public pressure can influence policy-makers to consider the test invalid for its intended purpose.

This paper discusses why and how test developers should consider what people think about the test. We argue that it is worth exploring the disparities between the intended purposes of the test and what people think the test actually measures - namely, the "face validity" of the test. Although face validity is considered a subjective judgment which is not indicative of the test scores' predictive or construct validities, we believe that it might offer a unique contribution to the development of a sound interpretive argument. The disparity between the intentions of the test developers and people's perceptions of the test can lead to various misinterpretations and misuses of test scores. Test developers must not only explain how test scores should be interpreted and used (Kane, 2006), but also why certain alternative interpretations and uses are inappropriate and may lead to undesirable consequences.

To support our claims, we present findings from a study about perceptions of the PET. Following the methodology outlined by Nevo (1985), we surveyed future, present and past PET examinees, faculty and admissions officers from institutes of higher education, and other public figures who deal with educational issues. We also surveyed the PET developers and other measurement experts. Apart from background information,

the data we collected included multiple-choice and open-ended questions about: (a) the relevance of the PET to its intended use, (b) the extent to which the PET measures abilities important for success in higher education, (c) the extent to which people agree with various claims for and against the PET, (d) ways that the PET scores are used that differ from their intended use, and (e) preferences among tools and methods for selecting students to higher education. The paper discusses the study findings, their implications, and how they were used to support proposed changes in the structure, content, administration and score interpretation of the PET.

Overall the results show the complex nature of opinions that the public holds about the test. In general, the public acknowledge the quality of the test design, administration and scoring. However, most people have negative views about the test's content and time constraints. Moreover, a large proportion of the public does not think the test is needed or that it successfully fulfills its intended purpose. These views vary somewhat between future students, current and past students, and academic staff. The results also show that some aspects of the test are particularly problematic in the public eyes. We show that some of these problems can be significantly reduced by future changes to the test design and administration.