

Abstract

REASONING WITH PSEUDOWORDS: HOW PROPERTIES OF NOVEL VERBAL STIMULI INFLUENCE ITEM DIFFICULTY AND LINGUISTIC-GROUP SCORE DIFFERENCES ON COGNITIVE ABILITY ASSESSMENTS

by

Paul Agnello

Advisor: Dr. Charles A. Scherbaum

Pseudowords (words that are not real but resemble real words in a language) have been used increasingly as a technique to reduce contamination due to construct-irrelevant variance in assessments of verbal fluid reasoning (Gf). However, despite pseudowords being researched heavily in other psychology sub-disciplines, they have received little attention in cognitive ability testing contexts. Thus, there has been an assumption that all pseudowords work equally and work equally well for all test-takers. The current research examined three objectives with the first being whether changes to the pseudoword properties of length and wordlikeness (how much a pseudoword resembles a typical or common word in English) led to changes in item difficulty on verbal Gf items. The second objective was whether boundary conditions existed such that changes to pseudoword properties would differentially impact two linguistic sub-groups of participants – those who have English as their dominant language and those who do not have English as their dominant language. The last objective was to index and explore performance on these verbal Gf items when pseudowords were replaced with real words. Hypotheses pertaining to predicting how pseudoword properties influenced item difficulty, how word type – pseudo or real impacted performance across linguistic sub-groups, and how linguistic sub-group status interacted with pseudoword properties were tested. Four sets of pseudowords were developed – short and wordlike, long and wordlike, short and un-wordlike, and long and un-wordlike, as well

as two sets of real words – short and wordlike, and long and un-wordlike. Sixteen verbal Gf items, adapted from the LSAT, were developed to accommodate the pseudowords or real words and explore these three objectives. While none of the hypotheses were statistically significant, the results did indicate further areas of exploration. Specifically, verbal Gf items were easier when they featured longer pseudowords and more difficult when they featured un-wordlike pseudowords. Additionally, while performance of English-non-dominant participants was fairly balanced across real and pseudoword sets, English-dominant participants performed better on items featuring real words. Similarly, linguistic status interacted with wordlikeness such that English-dominant participants featured a decrease in performance as pseudowords moved from wordlike to un-wordlike. A full discussion of the findings, their implications, limitations of the current study, and directions for future research are included.

Abstract

Despite females demonstrating excellent mathematics class performance and strong standardized test-taking abilities in non-mathematics domains, females have underperformed relative to males on standardized mathematics assessments such as the SAT. One counterintuitive finding is that females have performed worse than males on word/applied mathematics items despite typically excelling in verbal domains. The vast majority of research on the gender gap has used between-persons approaches which have been unable to fully accommodate female variability across mathematics sub-domains, item types, and individual test-taker characteristics impacting performance. Using data from the TIMSS 2007 mathematics exam, the present study used differential person functioning (DPF) analyses to document within-person differences in performance for female test-takers compared to males across computation and word/applied items when those items were matched on difficulty. Further, the present research sought to explain the variable nature of female mathematics performance as a function of the equally variable learning environment. Both the human capital and social capital present within the learning environment were examined as moderators of female achievement across item types via multilevel modeling. Results suggested that males performed significantly better on word/applied versus computation items compared to females. Human capital in terms of teacher years of experience significantly moderated the performance of female test-takers across item types while social capital and school location were found to be potential moderators in certain contexts. Implications and limitations of the present study are discussed and recommendations are made to guide theory, research, and policy regarding female mathematics achievement, teacher training, and teacher networks.

Researchers recognize that the current models of intelligence are insufficient at making causal connections between the intelligence measure and intelligent behavior. Different approaches to intelligence are under investigation to incorporate within current models of intelligence and include psychometric, neuropsychological, and cultural components. Currently there is a lack of research that incorporates both psychometric and neuropsychological intelligence measures in a predictive model of performance. The purpose of the current study is three-fold. The first objective is to test the predictive relationships of neuropsychological and psychometric intelligence batteries, an alternative psychometric intelligence assessment, and a personality measure in relationship to academic performance. The second objective is to examine racioethnic and gender subgroup mean differences on all predictors of performance. Subgroup mean differences, which can lead to adverse impact, have been found on a variety of verbal and nonverbal intelligence assessments (Hough, Oswald, & Ployhart, 2001). Research has demonstrated that performance differences are often moderated by the type of measure used which also raises concerns about the construct validity of psychometric intelligence assessments. The third objective of the research is to examine the construct validity of neuropsychological intelligence, traditional psychometric intelligence, and alternative psychometric intelligence. There is little empirical evidence which demonstrates that differences in cognitive functioning in the brain result in differences in scores on psychometric assessments. That is, there are few links (i.e., construct validity evidence) connecting cognitive functioning to intelligent performance on psychometric assessments. Hypotheses pertaining to prediction of different measures, subgroup mean differences, and statistical relationships among the intelligence measures were tested. The results indicate that the neuropsychological intelligence battery was the only significant predictor of academic performance. All intelligence measures exhibited subgroup mean differences, however they were smaller compared to what is typically reported in the literature. The Black/African American mean score on the neuropsychological battery was one-third of a standard deviation below the White/Caucasian mean score, and Hispanics demonstrated minimal mean score differences compared to White/Caucasians. Additionally, construct validity evidence emerged for the intelligence measures. A discussion of the findings including their implications is included.