International Journal of Educational Research 69 (2015) 71–87

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Development of an online test of problem solving ability that minimises the extraneous differential effects of language background and subsequent validation through a large-scale DIF analysis

Alvin Vista *, Esther Care, Patrick Griffin

Assessment Research Centre, Melbourne Graduate School of Education, The University of Melbourne, Level 8, 100 Leicester St, Parkville, 3010 Melbourne, VIC, Australia

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form 14 October 2014
Accepted 15 October 2014
Available online 1 December 2014

Keywords:
Differential item functioning
Problem solving
Language background
Reading comprehension
Test language

ABSTRACT

This paper presents a test of problem solving ability that is designed to minimise the differential effects of language background. Large-scale data are presented in order to determine the validity of this test through differential item functioning (DIF) analysis. The analysis was conducted on test data from 3527 students from 61 government schools across Victoria, Australia in 2010. The students have diverse linguistic backgrounds and a wide range of English language proficiency.

Using both 2PL IRT and logistic regression approaches it is shown that most items are not functioning differentially with respect to language background, providing evidence that the design of the instrument is validated by the data. The implications for future system-wide implementation of large scale interventions for Australia's linguistically diverse classrooms are outlined and discussed.

1. Introduction

This paper describes a test of problem solving ability that was designed to minimise the differential effects of language background. The developed test was validated by conducting a differential item functioning (DIF) analysis on a large and representative sample of government school students. The main importance of this tool can be viewed from the perspective that Australia has an increasing student population from non-English language backgrounds. Migrants from non-English speaking backgrounds are increasing, with almost half coming from Asia. The most recent Australian Bureau of Statistics data (2011 census) indicate that 53% of first-generation Australians now speak a language other than English at home, and 20% of second-generation Australians (Australian Bureau of Statistics, 2011). In Victoria in 2011, approximately 19.5% of the population were born overseas in non-main English-speaking countries while 23.1% spoke a language other than English at home (State of Victoria, 2013).

This demographic change has important implications for education in Victoria. One of the four major priorities of the Victorian Department of Education and Early Childhood Development's corporate plan is to "improve outcomes for

* Corresponding author. Tel.: +61413561957.
E-mail address: vistaa@unimelb.edu.au (A. Vista).

http://dx.doi.org/10.1016/j.ijer.2014.10.003
0883-0355/

disadvantaged young Victorians" (2008, p. 11). The implications for NESB students relate to the possible effects of cultural and language factors on this particular group's achievement levels. There is some evidence that NESB students perform at a lower level compared with English speaking background (ESB) students in state-wide and national achievement tests (Rothman, 2002).

With Australian classrooms becoming more diverse, it is timely to look at large scale data in order to describe the dynamic of how demographic factors may affect the measurement of domain-general and domain-specific student outcomes. National tests such as NAPLAN (Ministerial Council for Education, Early Childhood Development and Youth Affairs, 2009) have literacy and numeracy components, but do not include problem solving. This research allows measurement of a construct that may influence the numeracy component of academic performance while being relatively independent of schooling.

1.1. Measuring problem solving

The measurement of problem solving, similar to that of general reasoning ability, takes into account both inductive and deductive types of reasoning, as well as divergent and convergent thinking skills. While these skills are essential in a number of curriculum areas (e.g., numeracy), tests of general reasoning ability can be, and in fact are, usually constructed to be independent of specific learning domains. Problem solving tests measure general reasoning ability and creative thinking skills—domains that do not necessarily need any specialised knowledge in a learning area such as numeracy.

1.2. Test language and implications for non-native speakers

In cross-lingual assessments, dynamics of any association between the measures and the target latent variables are influenced by proficiency on the test language (TL) that is used. Especially for constructs that are theoretically language-independent, such as problem solving or mathematics, this issue between TL and student native language (or primary language, L1) can potentially confound the results. For example in the case of mathematics testing, has been shown that language skills (in particular, decoding and comprehension) have direct impact on math performance (Fuchs et al., 2006; Seethaler & Fuchs, 2006) or can be strongly associated with latent math ability (Hart, Petrill, Thompson, & Plomin, 2009) as the reading loads for these measures increase.

This effect of TL on math performance of non-native speakers is not always uniform because language loadings differ depending on the type of math skills being measured (e.g., arithmetic may have lower language loading than complex word problems) (Fuchs et al., 2005, 2006). There is evidence that skills in the TL impact the required reading skills which in turn overlap with the measure's loading on latent math ability (Hart et al., 2009). While this is more apparent in math, the same mechanism occurs in other domains whenever non-native languages are used for testing. Thus the TL has direct implications for non-native speakers of the TL, and it has been shown that "proficiency in a second language has theoretical and empirical relevance for interpreting the results of tests given in that language" (Alderman, 1981, p. 17). For non-native speakers, measures of constructs that are supposedly language-independent invariably become a test of language proficiency (Solano-Flores & Trumbull, 2003) and TL becomes a factor that contributes to measurement error while being in itself irrelevant to the construct being measured which confounds the valid interpretation of test results (Solano-Flores and Trumbull, 2003).

If the testing process takes into account too much of irrelevant factors, such as ability that is not relevant to the construct being measured, then there occurs what is known as construct–irrelevant variance, a major threat to validity where the statistical power of the predictor variable is diminished and results based on that measurement cannot be reliable (Messick, 1994). The language used in a test affects the results by introducing construct–irrelevant variance which, in turn, has implications for individuals from different linguistic backgrounds differently.

2. Method

2.1. The problem solving tests

The problem solving test items assess students' performance in key steps in the problem solving process including classification, application and reasoning, in situations that are both familiar and unfamiliar to the students. Items were developed such that they do not require specific academic knowledge. Although the items contain verbal loads, they require only basic English proficiency. The tests are mounted on the Assessment Research Centre Online Testing System (ARCOTS) which is designed to score and process the responses automatically using the psychometric information and item parameters contained in the item bank. The testing system and its items have been developed over several years by the Assessment Research Centre at the University of Melbourne (see Assessment Research Centre, 2011, 2012; Awwal, Pavlovic, & Robertson, 2010).

In developing the problem solving tests, a wide range of item difficulty (8 logits wide) and adequate discrimination ($>.10$) were used as initial parameters for inclusion in the item pool. Items fall into three broad groups based on the type of reasoning process involved: spatial, symbolic or verbal reasoning. They were mapped onto a continuous scale with overlapping items to link each level to the adjacent levels (see Supplementary Table S1).