Chapter 18
Mitigation of Test Bias in International, Cross-National Assessments of Higher-Order Thinking Skills

Raffaela Wolf
CAE, USA

Doris Zahner
CAE, USA

ABSTRACT

The assessment of higher-order skills in higher education has gained popularity internationally. In order to accurately measure the skills required for working in the 21st century, a shift in assessment strategies is required. More specifically, assessments that only require the recall of factual knowledge have been on the decline, whereas assessments that evoke higher-order cognitive skills are on the rise. The purpose of this chapter is to discuss and offer strategies for mitigating bias for a computer-administered performance-based assessment of higher-order skills. Strategies to abate the effects of bias are discussed within the test design and test implementation stages. A case study of a successful adaptation and translation of CAE's Collegiate Learning Assessment (CLA+) is presented to guide the discussion throughout the chapter.

INTRODUCTION

International academic institutions of higher education are under pressure to enhance the quality of instruction in order to equip the next generation's workforce with the skills necessary to meet the demands of careers evolving in the 21st century. Research suggests that employers seek individuals who are able to think critically and communicate effectively (Hart Research Associates, 2006). In order to meet the demands of today's world, a shift in assessment strategies is necessary to measure the skills now prized in a complex global environment. More specifically, assessments that only foster the recall of factual knowledge have been on the decline, whereas assessments that evoke

DOI: 10.4018/978-1-4666-9441-5.ch018

higher-order cognitive skills, such as analytic and quantitative reasoning, problem solving, and written communication are on the rise. CAE's Collegiate Learning Assessment (CLA+) is a computer-administered instrument that purports to measure these higher-order skills within the United States and internationally. In general, instruments that allow for cross-national comparisons of higher-order competencies in higher education are scarce, mainly due to an array of methodological challenges, such as differences in educational systems, socio-economic factors, and perceptions as to which constructs should be assessed (Blömeke, Zlatkin-Troitschanskaia, Kuhn, & Fege, 2013). From a more specific psychometric perspective, measuring competencies within an international framework poses challenges that pertain to test development, scoring, and the validity of score interpretations (R. K. Hambleton & Murphy, 1992). Different forms of bias are considered the main sources of in-equivalence in cross-national research (F. J. Van de Vijver, 1998; F. J. Van de Vijver & K. Leung, 1997). Van de Vijver and Leung (1997) identified three common forms of bias within an international context; which include construct bias, method bias, and item bias.

The purpose of this chapter is to discuss and offer strategies to minimize different sources of bias, such as construct bias, method bias, and item bias for a computer-administered performance based assessment of higher-order skills. Rather than viewing and/or treating each source of bias in isolation, a holistic approach that combines a mixture of qualitative and quantitate methodologies was employed with the hope of collecting evidence for valid cross-national score interpretations. The proposed strategies to abate the effects of bias are discussed within the test design (i.e., construct definition, item and rubric development) and test implementation (i.e., translation process, training of scorers, administration (delivery platform)) stages. A case study of a successful test adaptation and translation of CAE's CLA+ will be presented to guide the discussion throughout the chapter.

BACKGROUND

Increased globalization, among other factors, has created an increased interest in making cross-national comparisons on underlying constructs that the research instrument purports to measure. Occasionally, this necessitates translation of the instrument into a different language.

Due to differences in culture and language, among other differences in the population, it is evident that an examination of the degree to which the instrument measures the same construct across these cultural and language groups is a precursor to drawing valid score interpretations. In order to draw valid score inferences, it is assumed that individuals who earn the same observed score on these instruments have the same standing on the constructs underlying the measurement instrument. The evaluation of several criteria could aid in meeting the aforementioned assumption:

1. The construct measured exists across nations.
2. The construct is measured in the same manner across nations.
3. Items that are believed to be equivalent across nations are linguistically and statistically equivalent.
4. Similar scores across different adapted versions of the assessment reflect similar degrees of proficiency.

In order to address the points above and in an attempt to mitigate bias in the test translation and adaptation phase, quantitative methods should co-exist with both qualitative and cognitive methodologies to ensure quality control throughout the test adaptation process.