Article

Educational and Psychological Measurement
2014, Vol 74(1) 77–96

DOI: 10.1177/0013164413502669
epm.sagepub.com

Bias in Terms of Culture and a Method for Reducing It: An Eight-Country "Explanations of Unemployment Scale" Study

Kostas Mylonas[1], Adrian Furnham[2], and Country Collaborators[3]

Abstract

Several sources of bias can plague research data and individual assessment. When cultural groups are considered, across or even within countries, it is essential that the constructs assessed and evaluated are as free as possible from any source of bias and specifically from bias caused due to culturally specific characteristics. Employing the Explanations of Unemployment Scale (revised form) for a sample of 1,894 employed and unemployed adults across eight countries (the United States, the United Kingdom, Turkey, Spain, Romania, Poland, Greece, and Brazil), we applied a method based on individual differences multidimensional scaling and principal component analysis to detect item bias in terms of culture and try to eliminate this bias variance from the overall item variance so as to (a) avoid jeopardizing validity levels and (b) arrive at clearer and more meaningful dimensions after adjusting the raw scores by

[1]University of Athens, Athens, Greece
[2]University College London, University of London, London, UK
[3]Country Collaborators: William Divale, York College, City University of New York, USA; Cigdem Leblebici, Izmir University of Economics, Turkey; Sonia Gondim, Federal University of Bahia, Brazil; Angela Moniz, Centro Universitario Padre Anchieta, Brazil; Hector Grad, Autonomous University of Madrid, Spain; Jose Luis Alvaro, Complutense University of Madrid, Spain; Romeo Zeno Cretu, University of Bucharest, Romania; Ania Filus, University of Warsaw, Poland; and Pawel Boski, University of Warsaw, Poland

Corresponding Author:
Kostas Mylonas, Department of Psychology, University of Athens, Panepistimiopolis, Ilisia, 157 84, Athens, Greece.
Email: kmylonas@psych.uoa.gr

removing the bias part. The results supported our statistical–psychometric intervention as the structure computed for the unadjusted data was enhanced and clarified when the data were adjusted for bias in terms of culture. Finally, implications for individual assessment procedures are discussed, and a method for evaluating the relative impact of bias in terms of culture on the raw assessment scores is also presented.

Keywords

bias in terms of culture, individual differences Euclidean distance, explanations of unemployment, bias impact and elimination, individual assessment

Introduction

Testing processes are often plagued by serious sources of bias, such as administration errors, evaluation misjudgments, and incompatible or inappropriate norms. Other psychometric characteristics of the test employed, such as liability to response styles, structural deficiencies, and more (Byrne, 2008; Nunnally & Bernstein, 1994; van de Vijver & Leung, 1997), can also be a potential threat. One of these serious methodological and psychometric disadvantages is related to a person's special cultural characteristics and to the impact which construct nonequivalence levels (possibly induced by these characteristics) have on the assessment outcomes (van de Vijver, 2011). This fact has been completely disregarded in early cross-cultural research (Xu & Barnes, 2011). Sireci (2005, 2011) has supported that linguistic diversity poses threat and suggested bilingual designs for identifying items that do not function differently across languages. Such bias is now well known to methodologists and psychometricians working with cross-cultural data, with biased information being misleading and wrong (Hambleton & de Jong, 2003). Thus, to rule out such systematic bias, construct equivalence studies should be conducted *before* doing differential item functioning studies.

When dealing with bias detection and elimination, several terms require attention. With respect to the aims of the current study, attention was first directed to construct equivalence. Differences in scores between cultural groups can reflect valid differences in the construct measured *or* they can reflect—at least partially—measurement artifacts or bias. One major cause of bias in cross-cultural research is culture itself. Poortinga (1989), assuming ''it is meaningful to postulate the identity of psychological constructs cross-culturally'' (p. 738), has defined ''comparison scale'' as the identical scale cross-culturally formed by any hypothetical construct in terms of which a comparison is made; following this, he has discussed several ways of dealing with the artifacts in measurement caused by ''bias in terms of culture,'' as ''Which psychometric properties of data can be validly compared depends on which parameters of measurement scales can be taken as invariant across cultures'' (p. 740). In a satisfactory cross-cultural study, there is no variance left to be explained in terms of culture (Poortinga & van de Vijver, 1987). Thus, construct nonequivalence in cross-cultural studies can be mainly attributed to cultural variance, which paradoxically has to be reduced to null to derive cross-culturally comparable and meaningful structures.

Cultural characteristics, a basic notion in cross-cultural research, are difficult to define but they can be regarded as cultural identities with which a member of this culture abides. These may correlate with ecocultural indexes such as affluence and religion, education and population statistics, and even ecological facts for a country, as these can characterize an underlying cluster structure across seemingly different nations (Georgas & Berry, 1995). Thus, characteristics may be studied and may seem different across countries, but if the characteristics studied have common grounds across some of these units, then the number of cultures studied is less than the number of countries involved. Thus, cultural characteristics become even more difficult to define, as the ''cluster of nations'' method factor forming a homogeneous ''culture'' in a cross-cultural analysis is itself a part of the definition. Even more, such cultural characteristics can produce construct nonequivalence both across countries *and* across clusters of countries under study. Finally, different cultures may exist within the same country, as a culture may certainly be a subset of people within a country, since homogeneous subsets of people may possess specific characteristics that distinctly differentiate their way of cultural thinking from other homogeneous subsets of people within the same country (Mylonas, 2009a). For example, males and females within the same nation may possess and exhibit culturally different ways of interpersonal communication. Similar ''differences'' may exist under any kind of subset formation within a country-nation, such as occupation, place of residence, age groups, educational levels, and so on, their behaviors each time reflecting different cultural identities. Although this argument may initially sound contradictory to the Georgas and Berry (1995) arguments on clusters of nations, in fact it is a complementary one, as for the current argument, nations may ''conceal'' different cultures depending on financial, educational, religious, occupational characteristics, and even gender; in short, clusters of cultures within the same nation may exist.

Finally, the impact which cultural variations and characteristics of any type may have on a measure-item and on the assessed construct creates its differential functioning which can be briefly named ''bias in terms of culture.'' If bias in terms of culture can be efficiently treated (Poortinga, 1989; van de Vijver & Leung, 1997; van de Vijver & Poortinga, 2002), then construct equivalence can be sought as variance is set free to accommodate the true structure of the test under consideration avoiding bias side-effects, so structural equivalence can be much more easily studied.

Dealing With Bias in Terms of Culture

Several ways of dealing with bias in terms of culture have been proposed. These include and have mainly focused on item deletion, or in the case of cross-cultural research even in the deletion of whole countries from the overall sample. Removing item bias at the item level does not necessarily lead to equivalence, as it does not necessarily remove construct bias and bias in general. However, the question of bias in terms of culture elimination can, under conditions, be reduced to item bias with the biased item being treated as a disturbance at the item level that has to be removed (van de Vijver & Leung, 1997). Such removal of items, though, conceals potential danger as validity levels may be threatened (even content validity can be at stake; Byrne & van de Vijver, 2010).

There are statistical ways that have been proposed to detect and eliminate bias to achieve invariance across cultures. Some of these methods attempt to account for cultural variance by introducing confounding variables in the study design and then exclude variance/items by means of hierarchical regression models. Alternative ways include covariance structure analysis (CSA; Poortinga & van de Vijver, 1987) or structural equation modeling (SEM; Byrne, 2008; Byrne, Shavelson & Muthén, 1989; Byrne & van de Vijver, 2010). A separate note should be drawn with respect to SEM as it allows for equivalence testing and also encompasses flexible techniques (even for large-scale samples) so as to retain nonequivalent items in the analysis in the form of culture-specific indicators through the partial measurement condition (Byrne et al., 1989; van de Vijver, 2011).

Another method is to circumvent the cultural bias effect by controlling for external criteria, such as gender, age, ability, and other confounding variables specific in each culture, such as ecological indexes (Georgas & Berry, 1995); for intelligence testing, it has been shown—through partial correlation coefficients—that more than 50% of the items are biased (Valencia, Rankin, & Livingston, 1995). However, such an attempt may detect the source of bias but does not partial out unwanted variance. Other methods of bias detection have been extensively discussed by Sireci (2011). Van Hemert, Baerveldt, and Vermande (2001) have proposed that ''researchers who want to compare ethnic groups or groups with various levels of acculturation should carry out a study on the cultural bias of their items'' (p. 394), signifying the importance of detecting and possibly eliminating item bias that may be or contain cross-cultural bias or bias in terms of culture. However, their contribution in bias detection and elimination suggested replacement of items that show to be highly biased (following the traditional approach of item deletion) and did not allow for retaining useful parts of the biased items' variance. Still, the authors suggested that validity should be protected and that replacement of biased items should take place during a pilot study or by administering double in number items (with two versions per item) in any empirical-questionnaire study. Finally, a method of adjusting identified intercept differences when estimating latent means has been coined (Scholderer, Grunert, & Brunsø, 2005), a method similar in principle to our own.

In general, most of the aforementioned methods do not deal with individual assessment, as they are applied to samples. Even more, these methods are employed under the assumption that when deleting-replacing items to achieve invariance, this deletion-replacement remedies for bias overall and so this overall remedy is recursively transferred to the individual scores. The latter expectation though may be challenged, as deleting items may have serious impact on a scale's validity, so if we decide to delete items in cross-cultural (or cross-groups) research so as to achieve construct equivalence, we may end up with a scale that assesses a different construct than the original or intended one. This may have serious repercussions on individual psychological assessment apart from the empirical research ones. Byrne and van de Vijver (2010) have exemplified the process by referring to van de Vijver, Mylonas, Pavlopoulos, and Georgas (2006), where 7 out of the 18 items in all were deleted either beforehand or during equivalence testing. Although Byrne and van de Vijver conclude that the scale's validity was not harmed, this may not always be the case.

An alternative way of dealing with such bias might be to work as much as possible within the variance of each item. That is, instead of deleting unwanted items, it would be much better to retain all items with their variance as ‘‘free’’ as possible of unwanted bias in terms of culture. So, we might try to intervene at the item variance level, instead of totally discarding the item. Through empirical studies this would be expected to reveal correlations closer to the true value and thus achieve better or at least clearer factor structures than the ones achieved before the intervention. Even more, with respect to individual assessment and scores, the scale's validity—as initially described and supported in a standardization study or a similar project of test construction—would not be threatened at all. This way, the individual scores would be more meaningful with respect to the original theoretical framework, still free of as much bias in terms of culture as possible.

Aim of the Current Study

The aim of the present study is to further explore and support an existing method for detecting and possibly eliminating bias in terms of culture (Mylonas, 2009a, 2009b). This bias may initially appear in the form of item-bias with respect to the items of a scale that has been applied cross-culturally, but through the method, it should be shown that most—if not all—of this item's bias is in reality bias in terms of culture. Having shown this, we may proceed to the second stage of the method, that is, elimination or at least reduction of these bias levels, so as to discard as much unwanted variance as possible without eliminating any of the original scale items. For the method to be tested, in the current study we employed Furnham's Explanations of Unemployment Scale (EoU-Revised) as administered to eight country samples of employed and unemployed adults. For the modeling of the data before and after the intervention through our method, principal component analysis (PCA) designs were employed in our attempt to explain under both conditions the largest possible variance through the extracted dimensions (Merenda, 1997). Real factors (Kline, 1993) and not estimated ones from the data infinity of solutions should be compared across the two conditions at this stage; common factor models (e.g., via maximum likelihood methods) should be employed in future attempts when the structure of the scale per se would be under consideration. However, indicative maximum likelihood solutions were also computed as a preliminary common factor approach using the same bias-reduction method. Finally, we combined the PCA solutions along with CSA indexes to be able to monitor construct equivalence at all stages.