Journal of School Psychology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of School Psychology

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

Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology☆

Laura L. Pendergast [a,*], Nathaniel von der Embse [a], Stephen P. Kilgus [b], Katie R. Eklund [b]

[a] Temple University, United States
[b] University of Missouri, United States

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form 5 June 2016
Accepted 22 November 2016
Available online xxxx

Keywords:
Factor analysis
Invariance
Equivalence
Measurement
Bias
Diversity
SABRS
Categorical
School psychology

ABSTRACT

Evidence-based interventions (EBIs) have become a central component of school psychology research and practice, but EBIs are dependent upon the availability and use of evidence-based assessments (EBAs) with diverse student populations. Multi-group confirmatory factor analysis (MG-CFA) is an analytical tool that can be used to examine the validity and measurement equivalence/invariance of scores across diverse groups. The objective of this article is to provide a conceptual and procedural overview of categorical MG-CFA, as well as an illustrated example based on data from the *Social and Academic Behavior Risk Screener* (SABRS) – a tool designed for use in school-based interventions. This article serves as a non-technical primer on the topic of MG-CFA with ordinal (rating scale) data and does so through the framework of examining equivalence of measures used for EBIs within multi-tiered models – an understudied topic. To go along with the illustrated example, we have provided supplementary files that include sample data, Mplus input code, and an annotated guide for understanding the input code (http://dx.doi.org/10.1016/j.jsp.2016.11.002). Data needed to reproduce analyses in this article are available as supplemental materials (online only) in the Appendix of this article.

Over the past decade, school psychologists have been serving a more diverse student population than ever before due to demographic shifts in the United States, as well as the expansion of school psychology internationally (Jimerson, Stewart, Skokut, Cardenas, & Malone, 2009), and this trend is likely to continue (Ortiz, Flanagan, & Dynda, 2008). At the same time, evidence based interventions (EBIs), often implemented within multi-tiered models, have become a central focus of school psychology practice (NASP, 2010a), training (NASP, 2010b), and research (e.g., Glover & DiPerna, 2007). In a recent publication, the APA Division 16 Working Group on Translating Science to Practice concluded that, although there are many EBIs in school psychology, little is known about the effectiveness of most of EBIs among minority populations in diverse contexts and called for increased research on the use of EBIs across diverse groups (Forman et al., 2013). Although it is critical for researchers to examine the extent to which EBIs, and adaptations thereof, are effective across racial, ethnic, and other groups (Forman et al., 2013), there is a critical intermediary step related to assessment.

☆ Data needed to reproduce analyses in this article are available as supplemental materials (online only) in the Appendix of this article.

* Corresponding author.
E-mail address: laura.pendergast@temple.edu (L.L. Pendergast).
Action Editor: Michael Toland

http://dx.doi.org/10.1016/j.jsp.2016.11.002
0022-4405/

The success of EBI programs is dependent upon the availability and use of evidence-based assessments (EBAs; e.g., Mash & Hunsley, 2005; Youngstrom, Findling, Kogos-Youngstrom, & Calabrese, 2005). Assessment procedures used within EBIs must produce reliable scores that allow for valid inferences (Kratochwill & Stoiber, 2002). Moreover, according to the Standards for Educational and Psychological Testing, test users should select measures that are fair for all test takers. "A test that is fair within the meaning of the Standards reflects the same construct(s) for all test takers, and scores from it have the same meaning for all individuals in the intended population." (AERA, APA, NCME, 2014; p. 50). Given that the outcomes of EBIs must be evaluated using scores from EBAs that are culturally sensitive, fair, and valid, it logically follows that the examination of measurement equivalence (also known as measurement invariance; ME/I) of EBA tools across diverse groups is one necessary pre-requisite for examining the effectiveness of EBIs in multicultural contexts. Although the presence of ME/I by no means guarantees that test scores are culturally sensitive or fair, ME/I does provide one essential piece of evidence in support of comparable (structural) validity across groups. Multi-group confirmatory factor analysis (MG-CFA) is a useful tool for examining various levels of invariance (Chen, 2008; Gregorich, 2006). MG-CFA can be examined in many popular statistical packages (e.g., Mplus, R, EQS) and provides researchers and practitioners with important information about the comparability and usefulness of scores across cultural groups.

The purpose of this article is to (a) provide a primer on best practices in conducting MG-CFA[1] with ordinal, Likert-type, item response data (as most behavioral and social-emotional rating scales used in school psychology EBIs yield ordinal data), (b) illustrate the use of MG-CFA using ordinal data from the Social and Academic Behavior Risk Screener (SABRS; Kilgus, Chafouleas, & Riley-Tillman, 2013), and (c) emphasize the importance of ensuring that assessment tools produce valid scores across cultural groups. This article is a non-technical, conceptual and procedural overview for novices. Those interested in more technical and mathematically-oriented reviews are referred to other excellent resources (e.g., Kim & Yoon, 2011; Millsap, 2012; Millsap & Yun-Tein, 2004; Vandenberg & Lance, 2000). In this article, the explanations are non-mathematical with the intent of facilitating comprehension at an introductory level for non-statisticians. To our knowledge, this is the first manuscript to focus on ME/I or MG-CFA in the context of measures used for school-based, multi-tiered intervention models. Given that these intervention models are a burgeoning area of research in school psychology, discussion about evaluating the equivalence of measures used within these interventions may be overdue. Moreover, very few authors have examined the issue of categorical MG-CFA with ordinal (item response) data in a non-technical fashion that is accessible to non-statisticians and provided annotated syntax. This manuscript is divided into five sections (a) a review of ME/I and the historical context of the technique, (b) an overview that introduces key terms and concepts, (c) discussion of technical considerations when conducting MG-CFA analyses (e.g., necessary sample size, estimation techniques, etc.), (d) an illustration of the technique using a school-based measure that produces ordinal, Likert-type, item response data, and (e) a discussion of implications, complications, and advanced applications.

1. Section one: what is measurement equivalence/invariance?

1.1. Role of measurement

In psychology and education, researchers and practitioners regularly attempt to assess constructs that are believed to exist but cannot be observed directly (e.g., intelligence, mathematics achievement, depression, happiness). Because these constructs cannot be directly observed, the degree to which respondents possess an unobservable trait is typically inferred based on their responses to items (often responses to items on scales with Likert-type formats) that are believed to reflect the underlying construct. Measurement equivalence/invariance (ME/I) has been said to be achieved when the relationships between responses to items (indicators) and latent constructs are the same across groups (Drasgow & Kanfer, 1985). ME/I has often been discussed in the context of broader discussions regarding test bias and/or item bias.

1.2. Measurement equivalence/invariance and bias

1.2.1. Historical context for testing concerns

Historically, "test bias" has been among the most controversial and commonly misunderstood concepts in assessment and psychometrics. Over the past century, researchers, practitioners, and laypeople have raised concerns about "bias" in psychological tests across cultural groups. Many of these concerns are understandable given historical misuse of psychological test data. As one example, historically, some published studies have compared racial groups based on "intelligence" using scores from tests given in different nations *without* reasonably sufficient evidence of validity among members of the groups with which the tests were used (e.g., Porteus, 1937). At best, such research reflects poor science (inadequate design and insufficiently supported conclusions; see Linstrum, 2016); some have argued that it embodies scientific racism (e.g., Fairchild, 1991; Richards, 1997).

1.2.2. First generation of bias research

Given the troubled history of race and research in many fields (Dennis, 1995; Poortinga, 1995), it is not surprising that racial differences in mean scores on psychological and educational assessments (e.g., IQ tests, college entrance exams, such as the SATs, etc.) are often greeted with skepticism. In fact, the study of measurement invariance and bias in testing originated in response to concerns and skepticism regarding race and gender differences in mean scores on IQ tests and college entrance exams. Zumbo

[1] This primer focuses on the use of categorical MG-CFA with ordinal data. When the term MG-CFA is used, the reader can assume that the authors are referring to categorical MG-CFA with ordinal data except where otherwise noted.

Please cite this article as: Pendergast, L.L., et al., Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology, *Journal of School Psychology* (2016), http://dx.doi.org/10.1016/j.jsp.2016.11.002