



Contents lists available at ScienceDirect

Journal of School Psychology

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

Methods matter: A multi-trait multi-method analysis of student behavior[☆]



Faith G. Miller^{a,*}, Austin H. Johnson^b, Huihui Yu^c, Sandra M. Chafouleas^c,
D. Betsy McCoach^c, T. Chris Riley-Tillman^d, Gregory A. Fabiano^e, Megan E. Welsh^f

^a University of Minnesota, United States

^b University of California – Riverside, United States

^c University of Connecticut, United States

^d University of Missouri, United States

^e University at Buffalo, United States

^f University of California – Davis, United States

ARTICLE INFO

Action Editor: Andy Garbacz

Keywords:

Assessment

Behavior

Direct Behavior Rating

MTMM

Systematic direct observation

Validity

Rating scales

ABSTRACT

Reliable and valid data form the foundation for evidence-based practices, yet surprisingly few studies on school-based behavioral assessments have been conducted which implemented one of the most fundamental approaches to construct validation, the multitrait-multimethod matrix (MTMM). To this end, the current study examined the reliability and validity of data derived from three commonly utilized school-based behavioral assessment methods: Direct Behavior Rating – Single Item Scales, systematic direct observations, and behavior rating scales on three common constructs of interest: academically engaged, disruptive, and respectful behavior. Further, this study included data from different sources including student self-report, teacher report, and external observers. A total of 831 students in grades 3–8 and 129 teachers served as participants. Data were analyzed using bivariate correlations of the MTMM, as well as single and multi-level structural equation modeling. Results suggested the presence of strong methods effects for all the assessment methods utilized, as well as significant relations between constructs of interest. Implications for practice and future research are discussed.

1. Introduction

Students who develop emotional and behavioral disorders experience some of the poorest academic, social, and behavioral outcomes of any disability group (Bradley, Doolittle, & Bartolotta, 2008). As such, remediation of such difficulties is essential. The use of a prevention-focused approach that begins with using psychometrically sound assessment methods to support early identification, intervention, and problem solving for these students, followed by the delivery of evidence-based interventions and systems that actively support and enhance implementation of interventions, have the potential to substantially improve these students' trajectories. Reliable and valid assessment data form the foundation of such evidence-based practices; decisions and inferences are only as good as the available data. To this end, clear understanding of the strengths and limitations of our assessment data is essential. The

[☆] Preparation of this article was supported by funding provided by the Institute of Education Sciences, U.S. Department of Education (R324A110017: PI, Chafouleas). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education, and such endorsements should not be inferred.

* Corresponding author at: University of Minnesota, Department of Educational Psychology, 250 Education Sciences Building, Minneapolis, MN 55455, United States.

E-mail address: fgmiller@umn.edu (F.G. Miller).

<https://doi.org/10.1016/j.jsp.2018.01.002>

Received 14 October 2016; Received in revised form 7 December 2017; Accepted 24 January 2018

0022-4405/© 2018 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

purpose of this study was to examine the reliability and validity of school-based behavioral assessment data targeting three constructs (i.e., academic engagement, disruptive behavior, and respectful behavior) which may serve as critical indicators of student success in schools.

1.1. Remediating behavioral problems in educational settings

Given the substantial negative academic, behavioral, and social-emotional outcomes associated with problem behavior, researchers and practitioners have focused upon the identification of both prevention and intervention techniques targeted towards behavioral problems. The prevention framework that has been integrated into systems of service delivery across disciplines ([Institute of Medicine, 2012](#)) focuses broadly on avoiding problems through universal promotive strategies and early intervention with individuals who begin to exhibit risk for problems. This framework is reflected in formalized multi-tiered systems of support (MTSS) implemented by schools to promote student behavioral success through prevention and intervention, such as School-Wide Positive Behavior Interventions and Support (SWPBIS; [Sugai & Horner, 2002](#)) and Safe and Civil Schools ([Sprick, Booher, & Garrison, 2009](#)). Although the identification and implementation of effective practices to support the behavioral success of all students is critical, the decision-making process involved in determining student need, the identification of an intervention best suited to an individual student's presenting problem, and the ongoing monitoring of student responses to these interventions depend on the use of assessment data which demonstrate evidence for their validity and reliability in that specific decision-making context.

Behavioral assessment data are essential to understanding whether prevention and intervention services are warranted and effective ([Deno, 1986](#)). Accordingly, within MTSS, methods for behavioral data collection vary based upon the question of concern. At the whole-school level, data sources such as office discipline referrals (ODRs) and standardized screeners (e.g., [Walker, Cheney, Stage, & Blum, 2005](#)) may provide actionable information on levels of need and groups or individuals in need of additional support. At the small-group or individual-student level, where interventions are being implemented in order to remediate behavioral challenges before they become clinically significant or to treat specific behavioral concerns, data sources are often focused on progress monitoring for the purpose of evaluating student response to intervention ([Kazdin, 2005](#)). Relevant assessment methods include those that rely on a teacher or another professional's observation of discrete student behavior, such as systematic direct observation ([Suen & Ary, 1989](#)) and Direct Behavior Rating ([Chafouleas, Sanetti, Kilgus, & Maggin, 2012](#)), as well as rating scales which assess behavior from a larger retrospective period of time (e.g., [Gresham et al., 2010](#)). Given the central role that such data play in multi-tiered decision-making, the psychometric quality of behavioral assessment data is a critical consideration for researchers and practitioners who seek to improve behavioral outcomes for students.

1.2. Reliable and valid behavioral assessment data

The appropriateness of traditional psychometric standards and analyses as applied to behavioral assessment, and indeed the defining features of behavioral assessment itself, has been fraught with debate ([Silva, 1993](#)). Many researchers placed direct observation, accompanying the development of Applied Behavior Analysis and its related methodologies, in contrast with traditional methods of personality and cognitive assessment which had heretofore defined the psychological and educational assessment landscape. As described by [Silva \(1993\)](#), the introduction of direct observation as a cornerstone of assessment in a behavioral paradigm was accompanied by the proliferation of measures which were criticized as having little to no evidence for the quality of the resulting data.

As part of these concerns, extensive interest arose in determining precisely how to evaluate the quality of behavioral assessment tools, with significantly differing views presented across multiple decades of research. As [Silva \(1993\)](#) described, "It is rare when we feel we have 'incontrovertible' measures of anything, and in a philosophical sense incontrovertible standards simply do not exist. No matter how fine-grained the measurement, we always only approximate reality" (p. 498). With that history acknowledged, it is worth revisiting two foundational concepts within psychometrics, reliability and validity, in order to provide context for both and define our position for each within behavioral assessment intended for use in decision-making within multi-tiered frameworks.

1.2.1. Reliability and validity

Reliability, broadly defined as a metric of consistency in scores along some dimension of interest, can be gauged using a variety of methods ([American Educational Research Association \[AERA\], American Psychological Association \[APA\], & National Council on Measurement in Education \[NCME\], 2014](#)). Given the distinct assumptions made by different assessment methods, the particular approach used to estimate a reliability coefficient should vary by assessment format. For example, when scores are repeatedly collected over time and aggregated, one might use an intraclass correlation to estimate the consistency of the collection of scores. This reflects an assumption that this aggregated score should be derived from individual observations which reflect systematic change or consistency in an object of measurement, rather than unsystematic change or error. In contrast, interrater reliability estimates such as the agreement index or the kappa coefficient might be used for observational assessment among multiple raters, when consistency within observations across raters is the topic of interest. Finally, when methods involve both observation and multiple timepoints, multiple reliability metrics can be used.

Perhaps the most important notion with regard to reliability, however, is the key role that it plays in validation. Validity, or the extent to which it is appropriate to use a test in a particular way, is at the heart of most psychometric arguments, and scores cannot be valid if they are not reliable because unreliable scores will not yield a consistent signal about students' behavioral status. However, reliability is only the first step. To establish that it is appropriate to use an assessment for screening or progress monitoring,

Download English Version:

<https://daneshyari.com/en/article/6843729>

Download Persian Version:

<https://daneshyari.com/article/6843729>

[Daneshyari.com](https://daneshyari.com)