Journal of Applied Developmental Psychology xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

A continuous improvement approach to social and emotional competency measurement

Laura A. Davidson [a,*], Marisa K. Crowder [a,b], Rachel A. Gordon [d], Celene E. Domitrovich [c,e], Randal D. Brown [a,b], Benjamin I. Hayes [a]

[a] Washoe County School District, Office of Accountability, 425 E. 9th St., Reno, NV 89521, United States
[b] University of Nevada, Reno, Interdisciplinary Social Psychology PhD Program, 1664 N. Virginia St. Mailstop 1300, Reno, NV 89557, United States
[c] CASEL: The Collaborative for Academic, Social, and Emotional Learning, 815 West Van Buren St., Suite 210, Chicago, IL 60607, United States
[d] University of Illinois at Chicago, Institute of Government and Public Affairs, 815 West Van Buren St., Chicago, IL 60607, United States
[e] Georgetown University, Department of Psychiatry, MedStar Georgetown University Hospital, 3800 Reservoir Road, NW, Washington, DC 20007, United States

ARTICLE INFO

Article history:
Received 1 October 2016
Received in revised form 13 February 2017
Accepted 9 March 2017
Available online xxxx

Keywords:
Researcher-practitioner partnership
Social emotional competency
Item response theory
Continuous measure improvement
Social emotional learning

ABSTRACT

A four-year research project by a researcher-practitioner partnership to develop a self-report measure of student social and emotional competency to identify at-risk students and guide practice is described. A continuous measure improvement approach facilitated work towards this goal. Items developed collaboratively by the partnership team were administered annually to all 5th, 6th, 8th, and 11th graders in WCSD in the context of their online school climate survey. Despite strong construct validity, initial Rasch analysis revealed a substantial ceiling effect inhibiting assessment of students at the mid-to-high range of social and emotional ability. Student focus groups informed by a latent class analysis were conducted and expert practitioners and scholars refined the items. The process resulted in an improved measure as well as more consistent district-wide survey administration. Strengths and challenges of the scale development process and data use strategies are discussed along with recommendations for future assessment development efforts.

In recent years, increased attention has been paid to the importance of students' social and emotional competencies (SECs), or the knowledge, skills, and attitudes needed to be personally and socially competent. The enhanced focus is the result of research documenting that students with SEC perform better in school, are more likely to stay in school and graduate, and function at higher levels in their adult lives than students without SEC (e.g., Farrington et al., 2012; Jones, Greenberg, & Crowley, 2015; Valiente, Swanson, & Eisenberg, 2012; see also Schamberg et al., 2017). A lack of SEC also appears to be a risk factor for poor outcomes. Students who enter school with lower SECs have been shown to fall behind their peers in early elementary school and are at greater risk in adolescence for social adjustment problems, academic failure, and drop out (Arsenio, Adams, & Gold, 2009; Barry & Reschly, 2012; Domitrovich, Durlak, Staley, & Weissberg, 2017).

Given the importance of SEC, particularly for school success, it is surprising that relatively less attention has been devoted to assessment of these competencies. McKown (2016) provides several recommendations for ensuring that measures intended to assess SECs in educational contexts are high quality and useful. These include, among others, the need for SEC assessments to be: 1) conducted on a large scale without the need for trained clinicians or researchers, 2) based on strong theoretical models, 3) informed by educators so that they are practical and solve "real world" problems that teachers care about, and 4) able to assess a range of dimensions that can develop a comprehensive picture of a student's social and emotional needs and strengths.

In line with these recommendations, current Standards for Educational and Psychological Testing call for an approach to measure development that views reliability and validity as fluid properties that vary across populations, locations, and time, rather than fixed traits of instruments (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The Standards also encourage test adopters to weigh reliability and validity evidence with the intended use of the measure in mind, and emphasize the importance of considering the applicability of existing evidence for local contexts, the accumulation of new evidence, where needed, and the periodic revision of tests. This kind of iterative test development and refinement process has been common for standardized tests that assess academic constructs, which are often developed by testing companies who regularly re-examine and refresh item pools and accumulate evidence through multiple

* Corresponding author.
E-mail addresses: LDavidson@WashoeSchools.net (L.A. Davidson), mcrowder@unr.edu (M.K. Crowder), Ragordon@uic.edu (R.A. Gordon), cdomitrovich@casel.org (C.E. Domitrovich), randalb@unr.edu (R.D. Brown), BlHayes@washoeschools.net (B.I. Hayes).

http://dx.doi.org/10.1016/j.appdev.2017.03.002
0193-3973/

strategies including large norming samples as well as cognitive interviews with test takers and content reviews by experts. Although testing companies do not always implement this continuous improvement approach completely and successfully, scrutiny from stakeholders and competition for large contracts has encouraged its intentional adoption throughout the test development, delivery, and revision process (Wild & Ramaswamy, 2008).

In the area of SECs, greater attention is spawning new assessments, and existing measures are sometimes developed with, or disseminated by, testing and publishing companies. Yet, historically, SEC measures have also frequently been developed by independent scholars or research teams, sometimes reflecting substantial content knowledge, but not the continuous measure improvement approach discussed above. Many developers of these measures have aimed for a fixed and relatively small list of items, with scoring based on simple item sums or averages, and revisions avoided to maintain cross-time comparability (see Denham, 2016 for a description of several established SEC measures). As a result, assessments are rarely adapted to meet the needs of local contexts. Past reliance on classical test theory approaches has also focused on test-level information for reliability and validity evidence (e.g., Cronbach's alpha). In contrast, the IRT approaches more common in testing firms offer detailed item-level statistics, support large item banks, and readily link old and new item sets through common items (Gordon, 2014, 2015). Leveraging these strengths of IRT can facilitate continuous improvement, including adjusting to school contexts where districts regularly re-administer surveys to students and may want to tailor items to local needs (including evolving student demographics, applicability across grade levels and to state-, district- or school-standards and pedagogy, and desire to present students with fresh items rather than the same small set year after year).

The project described in this paper applied the continuous measure improvement approach to the development of a student self-report rating of social and emotional competence. Implementation of our approach relied on a researcher-practitioner partnership, which bridged the psychometric expertise of academic scholars with the on-the-ground research and practice expertise in the school district. This project used a four-year, mixed-method, iterative approach to identify and address the key measurement challenges of the instrument, focusing on the intended goal of the partnership: to assess a large population of students with a range of SEC ability so that early intervention with students who need instructional support can occur. Rasch-based analyses of student survey items provided item-level statistics, latent class analyses identified groups of students that contributed to an identified ceiling effect, and focus groups explored reasons for the ceiling effect and helped refine items and the survey environment. This approach also offered local validity evidence by developing and refining items in relation to school district standards for SEC instructional practice and with input from teachers and students. The project also emphasized a continuous improvement approach, using item-level statistics based on analyses of each year's survey to iteratively improve the item bank, ensure that items covered the full range of each construct, and allow selection of item subsets to be used at particular grade levels and in repeated administrations.

The project resulted in the development of a bank of 138 student self-report items that assess eight subdimensions of SECs. From this larger bank of items, two subsets (which we refer to as instruments in order to facilitate communication with internal and external audiences) were also created. The 138–item bank allows for the flexible creation of subsets for various uses, offers extensive examples of indicators aligned to the five SEC clusters defined by the Collaborative for Academic, Social, and Emotional Learning (CASEL), and has been useful for training purposes and cross-walking to learning standards. The first instrument includes 40 items that the partnership selected because they covered all subdimensions of interest, were well-aligned with district standards for SEC instructional practice, and performed well in iterative administrations with Washoe County School District (WCSD) students. The second is a 17-item instrument that the partnership felt would be useful as a short-form assessment of global SEC for when districts needed fewer than 40 items due to concerns about time and burden. This short-form measure is currently in use by several districts in standalone form or as part of a larger climate survey (see Washoe County School District, & Collaborative for Academic, Social, and Emotional Learning, 2016). The full item bank and both the 17- and 40-item instruments are free, open-source and available for other districts to use and adapt for their measurement needs. This paper focuses on two subdimensions from the larger item bank, Relationship Skills (RS) and Self-Management of Emotions (SME), to illustrate the collaborative process undertaken by the partnership to address the ceiling effect found in the original measure and to produce an assessment of student social and emotional competence with evidence of local reliability and validity.

1. The WCSD-CASEL research-practice partnership project

CASEL and the NoVo Foundation developed the Collaborating Districts Initiative (CDI) to build the capacity of eight large urban school districts, including WCSD, to systematically provide support for social and emotional learning (SEL) to all preK-12 students and the adults who served them. SEL is the process through which students acquire and apply the knowledge, skills, and attitudes needed to be personally and socially competent. The goals of the CDI are to demonstrate that a district can implement SEL with fidelity, at scale across a system, as well as to collaboratively develop and refine practical tools that promote the effective implementation and assessment of SEL, and strengthen the research base related to this work. All districts in the CDI participated in a national evaluation conducted by the American Institutes for Research (AIR) which included the collection of student self-report ratings of social and emotional competence (American Institutes for Research & Collaborative for Academic, Social, and Emotional Learning, 2013). It was these ratings that were iteratively refined in the current project.

In 2009, WCSD had developed an Early Warning Risk Index (EWRI) to help identify and intervene with students at risk for not graduating based on factors that paralleled Balfanz's original academic risk prediction model (Balfanz, Herzog, & Mac Iver, 2007; Balfanz & Byrnes, 2006), including credit deficiencies, test scores, absenteeism, suspensions, transiency, and retention. WCSD conducted a longitudinal analysis of the EWRI's effectiveness in predicting graduation rates for a sample of students who were 9th graders in the district in 2009. Results showed that only one-third of all “High Risk” 9th grade students beat the odds and graduated four years later, compared with 79% of students identified as “No Risk” in the 9th grade. Although impressed at its predictive utility, WCSD developed an interest in the 34% of high risk students who had some resilient characteristics unidentified by the model that allowed them to overcome substantial academic obstacles in the 9th grade and graduate four years later.

As a result of participating in the CDI, WCSD had collected the self-report SEC ratings mentioned above and wanted to explore whether SEC served as a protective factor that contributed to student resilience. Ultimately, WCSD intended to include the measure as part of its interactive online data reports which include the EWRI scores alongside academic and behavioral data to allow for improved educational decision-making. CASEL intended to use the measure to disseminate nationally with guidance to districts interested in using the assessment as part of their SEL measurement systems. Given these intended uses, both organizations wanted to deepen and broaden the evidence base about the measure.

The district data on student SEC, collected through its participation in the CDI evaluation, showed the subdimensions of the measure had moderate reliabilities ($\alpha = 0.68$ to 0.74), but also substantial ceiling effects. For example, 18% of WCSD students rated themselves at the highest level on all four items (*Very True for Me*) in the Social Awareness subdimension, and 14% rated themselves at the highest level on all seven items of the Self-Management subdimension. High self-ratings

Please cite this article as: Davidson, L.A., et al., A continuous improvement approach to social and emotional competency measurement, *Journal of Applied Developmental Psychology* (2017), http://dx.doi.org/10.1016/j.appdev.2017.03.002