EdInstruments

Considerations for Use of Assessments of Social-Emotional Competence[1]

As part of the larger library of instruments, EdInstruments catalogues assessments of social-emotional competence (and related scales) for use in research and practice. To guide potential users, we define social-emotional competence, highlight general psychometric properties to consider, and underscore specific considerations for use of such instruments by school practitioners. While we do not endorse individual instruments, we provide examples of instruments for various uses. Users can navigate our website to explore additional options.

Social-Emotional Competence

Social-emotional competence has been defined in a variety of ways. Most often, these skills are referred to as soft skills, non-cognitive skills, or 21st century skills and refer to competencies beyond traditional academic knowledge and skills. A range of related frameworks have been developed over the years. While we acknowledge the complexity in this area, EdInstruments adopted one of the most widely-utilized frameworks from CASEL, which includes Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision-Making. We highlight CASEL's definitions of each domain below. According to CASEL, social and emotional learning

> is the process through which all young people and adults acquire and apply the knowledge, skills, and attitudes to develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make responsible and caring decisions.

Self-Awareness: "the abilities to understand one's own emotions, thoughts, and values and how they influence behavior across contexts. This includes capacities to recognize one's strengths and limitations with a well-grounded sense of confidence and purpose."

[1] We thank Stephanie M. Jones, Deborah Moroney, Michael Rodriguez, Jim Soland, and Judy Wiener for their helpful feedback and expertise. Their participation in this work does not signify endorsement of any individual instrument/assessment. Any errors or omissions are our own.

Self-management: “the abilities to manage one’s emotions, thoughts, and behaviors effectively in different situations and to achieve goals and aspirations. This includes the capacities to delay gratification, manage stress, and feel motivation and agency to accomplish personal and collective goals.”

Social awareness: “the abilities to understand the perspectives of and empathize with others, including those from diverse backgrounds, cultures, and contexts. This includes the capacities to feel compassion for others, understand broader historical and social norms for behavior in different settings, and recognize family, school, and community resources and supports.”

Relationship skills: “the abilities to establish and maintain healthy and supportive relationships and to effectively navigate settings with diverse individuals and groups. This includes the capacities to communicate clearly, listen actively, cooperate, work collaboratively to problem solve and negotiate conflict constructively, navigate settings with differing social and cultural demands and opportunities, provide leadership, and seek or offer help when needed.”

Responsible decision-making: “the abilities to make caring and constructive choices about personal behavior and social interactions across diverse situations. This includes the capacities to consider ethical standards and safety concerns, and to evaluate the benefits and consequences of various actions for personal, social, and collective well-being.”

General Psychometric Considerations

Defining Validity

Validity is a topic of great importance that is also often misunderstood. Validity is the extent to which theory and evidence support the intended interpretations of scores for proposed uses. At a high level, instruments can only be supported by validity evidence for particular interpretations and uses with particular populations. Even if prior research supports the use of a given measure, adapting it for a new group of students or different purpose could render that research moot, or at least make it less relevant. In addition, consequential evidence is also important: whatever the technical qualities of the measure, ensuring that it leads to beneficial outcomes for students, or at least does no harm, should be paramount. See, for example, this recent paper on validity and its importance. When evaluating validity evidence of instruments under consideration, the relevant evidence should directly support your interpretations and uses.

Intended Interpretations and Uses

Users must consider the intended purpose of the instrument (i.e., for whom the instrument is intended to be used, the proposed purposes, and contexts in which the instrument is used) and whether this matches the intended use of the eventual user. For example, what is the intended age range? Is there an intended (sub-)population under consideration? Is the definition of the construct being measured consistent with your intended interpretation and support your use?

Content

An important source of evidence concerns the content of the questions comprising each measure. The greatest threats to validity include weak or limited content or construct coverage and

construct-irrelevant variance. Is the instrument content supported by theory and prior research? Was there expert review to support the content coverage? Is there evidence that instrument items were reviewed by experts for sensitivity to gender, race/ethnicity, linguistic complexity, culture, or other characteristics relevant to your intended uses?

Internal Structure
Is there research confirming that the construct is unidimensional (if that is the claim) or that it matches some other factor structure? Do all items contribute to the total score (evidence from item-total correlations or item-factor loadings)?

Consistency across Groups
Is there evidence that the measurement model parameters are consistent across groups, i.e., that there is no bias across gender, English learner (EL) versus non-EL students, students of different race or ethnicity? Evidence should, at a minimum, include differential item functioning (DIF) analyses, or better yet, measurement invariance or multigroup confirmatory factor analyses.

Consistency over Time
If the measure is intended for use longitudinally, such as in growth models, pre/post interventions, or value-added models, is there evidence that the measurement model parameters are consistent over time (longitudinal invariance)? Is there evidence of test-retest reliability?

Associations with Other Important Variables
To help support score interpretation, scores should be associated with measures of similar constructs and with theoretically relevant outcomes, such as school academic performance, educational aspirations, school engagement and behavior, risk-avoidance and resilient behaviors. Is there evidence that scores from the instrument are associated (correlated) with other important variables, such as other educational outcomes or behaviors? Do scores predict future behaviors?

Validity of the Scoring Approach
What is the recommended scoring model? Is there evidence that, if a sum score is recommended, score reliability is strong (at least .70 at a minimum, greater than .80 for group-level decisions, greater than .90 for individual level decisions)? Evidence suggests that sum scores make large, often untenable assumptions that bias inferences (McNeish & Wolf, 2020). Those assumptions only become more problematic when using scores over time (Kuhfeld & Soland, 2020; Soland & Kuhfeld, 2020). A stronger scoring model is based on item response theory (IRT), but this also requires stronger assumptions. Are there IRT/CFA item parameters from a larger population that can be used to score the survey? Is there a raw-score to scaled-score conversion table supported with empirical evidence? Calibrated item parameters could help avoid much of the bias that comes with scoring these surveys using samples from small interventions.

Specific Considerations for School Practitioners

The now sunsetted Funders' Collaborative for Innovative Measurement supported multiple interrelated projects to describe, inventory, and elevate assessments of "hard to measure" outcomes. This investment resulted in two practitioner guides that presented curated and searchable lists of assessments of social and emotional competencies: the CASEL Assessment

Guide and the RAND Online Measure Repository. AIR developed a collated guide in 2015, which was updated in 2019 with additional measure types (e.g., settings, adults). Additionally, CASEL and their partners convened a working group, Establishing Practical Social-Emotional Competence Assessments of Preschool to High School Students project, and an adjacent working group convened by AIR and Dale Blyth, the National Practitioner Advisory Group on Using Data to Inspire SEL Practice (NPAG). The NPAG produced a statement that contains 10 beliefs on SEL measurement from the perspective of practitioners.

Beyond these 10 beliefs, we also recognize additional issues that are critical to practitioner use, such as access (i.e., cost, ease of use), transparency (i.e., how clear is the purpose and audience for the instrument), fit (i.e., intended for practice use), and practicality (i.e., time for administration).

In addition, the BUROS Center for Testing developed an SEL Assessment Technical Guidebook for educators. This guidebook provides additional guidance to identify and select SEL instruments, evaluate measurement quality of instruments, and address issues of accessibility and fairness.

Recognizing these considerations, school practitioners may consider the following assessments for use in schools. For educators looking for a free, comprehensive assessment of social-emotional competence, the Washoe County School District Social and Emotional Competency Assessments (WCSD-SECAs) may be a good option to consider. Other widely used options include the Panorama Social-Emotional Learning Student Survey, SELweb, and the suite of Devereux Student Strengths Assessments. Individual constructs can also be searched on EdInstruments using keywords (e.g., grit, growth mindset, or motivation).

Diagnostics and Screeners

Furthermore, school practitioners work within schools with wide adoption of multitiered systems of support and other advancements in tiered interventions (which aim to keep most students in the general population while providing more intensive interventions and supports for those who would benefit from them). These systems require assessments that allow educators to explore and define students' needs. These assessments tend to require specific qualifications, such as a special education credential, qualification as a school counselor or psychologist, or are used by a counselor or educator with a clinical background. Some common examples of these assessments include the Beck Youth Inventories – Second Edition (BYI-2), the Behavior Assessment System for Children – Third Edition (BASC-3), and the School Social Behavior Scales – Second Edition (SSBS-2).

Concluding Remarks

To reiterate, measurement of social-emotional competence is still a developing field. Both researchers and practitioners should carefully consider issues of validity when choosing an instrument, as we highlight above.

In practice, measures of social-emotional competence and related diagnostic instruments and screeners should be used to inform teaching and learning as well as student support. Instruments requiring some level of clinical qualification should not be used otherwise.

Finally, instrument developers need to do better to clearly and accessibly present the purpose, burden, rigor, and intended use of the instrument up front – for the benefit of the field at large. We urge users to visit the external resources we highlight to assist them with their selection.

Additional References

Kuhfeld, M., & Soland, J. (2020). Avoiding bias from sum scores in growth estimates: An examination of IRT-based approaches to scoring longitudinal survey responses. *Psychological Methods*.

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 1-19.

Soland, J., Hamilton, L. S., & Stecher, B. M. (2014). Measuring 21st century competencies: Guidance for educators. Santa Monica, CA: RAND Corporation.

Soland, J., & Kuhfeld, M. (2020). Do response styles affect estimates of growth on social-emotional constructs? Evidence from four years of longitudinal survey scores. Multivariate Behavioral Research, 1-21.