Challenges and opportunities in the applied assessment of student social and emotional learning

Author: Clark McKown (ORCID: 0000-0001-9694-1179)

Publication Year: 2019

Journal: Educational Psychologist

Publisher: https://www.tandfonline.com/

Grant #: R305A160053

Educational Psychologist

ISSN: 0046-1520 (Print) 1532-6985 (Online) Journal homepage: https://www.tandfonline.com/loi/hedp20

Challenges and Opportunities in the Applied Assessment of Student Social and Emotional Learning

Clark McKown

To cite this article: Clark McKown (2019) Challenges and Opportunities in the Applied Assessment of Student Social and Emotional Learning, Educational Psychologist, 54:3, 205-221, DOI: 10.1080/00461520.2019.1614446

To link to this article: https://doi.org/10.1080/00461520.2019.1614446

Published online: 11 Jun 2019.

Submit your article to this journal

Article views: 123

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hedp20

EDUCATIONAL PSYCHOLOGIST, *54*(3), 205–221, 2019
Copyright © 2019 Division 15, American Psychological Association
ISSN: 0046-1520 print / 1532-6985 online
DOI: 10.1080/00461520.2019.1614446

Challenges and Opportunities in the Applied Assessment of Student Social and Emotional Learning

Clark McKown

Department of Behavioral Sciences, Rush University Medical Center

Interest in school-based strategies to support student social and emotional learning (SEL) is strong. Although SEL policies and programs designed to support the development of student competencies have advanced significantly, less work has been done to develop methods of assessing student social and emotional competence. This article briefly reviews developments in the field of social and emotional competence assessment and examines challenges and opportunities in their applied use, including (a) balancing the priorities of assessment developers and educators; (b) ensuring that the inferences and decisions made from SEL assessment scores are supported by evidence of the assessment's psychometric merit; (c) establishing conditions for SEL assessment and data use that maximize benefit while mitigating risks; (d) coordinating standards, assessment, programs, and professional learning; and (e) balancing highly focused assessments that by design do not vary in content or format, and the varied cultural contexts in which they may be used.

Universally administered student social and emotional competence assessments have the potential to support teaching and learning. Their use for this purpose is comparatively new, however, and it is important to take stock of possibilities and limitations. To that end, this article first provides a brief overview of the field of student social and emotional competence assessment—reviewing what social and emotional learning (SEL) is, why it matters, and what kinds of social and emotional competence assessments are currently available. It then explores five challenges and opportunities facing the field that I hypothesize will affect the likelihood that universal student social and emotional competence assessments will support teaching and learning. Those challenges and opportunities include (a) balancing the priorities of assessment developers and educators; (b) ensuring that the inferences and decisions made from SEL assessment scores are supported by evidence of the assessment's psychometric merit; (c) establishing conditions for SEL assessment and data use that maximize benefit while mitigating risks; (d) coordinating standards, programs, assessment, and professional learning; and (e) balancing highly focused assessments that, by design, do not vary in content or format, and the varied cultural contexts in which they may be used. The article describes each challenge and its implications for practice and offers recommendations for the field to move forward constructively.

Correspondence should be addressed to Clark McKown, Department of Behavioral Sciences, Rush University Medical Center, RNBC 4711 Golf Road, Suite 1100, Skokie, IL 60076. E-mail: Clark_A_McKown@rush.edu

OVERVIEW OF THE FIELD OF STUDENT SOCIAL AND EMOTIONAL COMPETENCE ASSESSMENT

What SEL Is and Why it Matters

The Collaborative for Academic Social and Emotional Learning (CASEL) has developed a widely cited and highly influential model, which defines SEL as "the process through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions" (CASEL, n.d.-b, para. 1). The knowledge, attitudes, and skills in this definition, which we refer to here as *social and emotional competencies*, span five broad areas: self-awareness, social awareness, self-management,

relationship skills, and responsible decision-making (CASEL, n.d.-a). The CASEL definition is widely used, is reflected in state SEL standards, and includes the competencies that are commonly the focus of instruction in SEL curricula and programs. Because of its ubiquity and influence, this article uses the CASEL model as the basis for describing opportunities and challenges in the applied assessment of social and emotional competencies.

However it is defined, there is substantial evidence supporting the conclusion that the broad range of social and emotional competencies is associated with children's success in school and with a wide range of concurrent and later life outcomes (see Durlak, Domitrovich, Weissberg, & Gullotta, 2015; Jones & Doolittle, 2017). As a result of its evident benefits to student success, a large number of universal school-based SEL programs have been developed and field tested. Meta-analyses have reported that when well implemented, these programs benefit students socially, behaviorally, and academically (Corcoran, Cheung, Kim, & Xie, 2018; Durlak Weissberg, Dymnicki, Taylor, & Schellinger, 2011) and that those benefits persist over time (Taylor, Oberle, Durlak, & Weissberg, 2017).

Because of strong evidence of the consequences of social and emotional competencies and that these competencies can be taught, it is not surprising that SEL has become a prominent topic in the national conversation about education from pre-K to Grade 12 in the United States. This is reflected in prominent developments in policy and funding priorities. For example, the Aspen Institute's National Commission on Social, Emotional, and Academic Development (2019) recently released a report to the nation with specific action steps to support student social and emotional development. At the same time, a growing number of states are incorporating social and emotional learning into their educational standards (Dusenbury, Dermody, & Weissberg, 2018). In addition, an increasing amount of funding is being committed to SEL, too. This is reflected in school budgets: Recent estimates suggest that public schools invest more than $600 million per year on SEL products and programs (Krachman & LaRocca, 2018). It is also reflected in philanthropic giving: The Allstate Foundation recently pledged $45 million to reach 25% of U.S. students with social and emotional programing (Molnar, 2018). In this policy and funding context, private and nonprofit organizations are launching or bringing to scale innovative methods of supporting educators in their work to teach social and emotional skills.

Universal Student Social and Emotional Competence Assessment

Among the developments in the field of SEL, one important consideration has been comparatively neglected: Relatively little work has focused on the assessment of student social and emotional competencies, referred in this article to as "SEL assessment." This appears to be changing. In addition to the general investments in SEL described previously, there is increased investment in SEL assessment specifically. For example, funding agencies such as the Institute of Education Sciences have supported efforts to develop and field-test social and behavioral assessments, including those designed to measure social and emotional skills. For the past 2 years, a working group of scholars and practitioners has engaged in several activities to build the field of SEL assessment (http://measuringsel.casel.org/). Private capital is beginning to support the growth of SEL assessment (Wan, 2017). Partly as a result of these investments, examples of widespread and innovative SEL assessment are beginning to emerge (McKown & Taylor, 2018).

These developments suggest that there is substantial and growing activity among assessment developers and funders and that there is a meaningful appetite for universal SEL assessment from preschool through high school. If the increased activity level and interest are any indication, educators will soon have a large number of SEL assessments from which to choose. Anticipating growth in the range of SEL assessments, this article examines five interrelated opportunities and challenges.

The focus of this article is on the assessment of student social and emotional competence. These are assessments specifically designed to measure the social and emotional knowledge, skills, and dispositions that children engage during interpersonal interactions and participation in school and community life (McKown, 2017b). I focus on student social and emotional competence because these outcomes, alongside academic competencies, are increasingly the focus of standards, programs, and classroom practices. As a result, well-developed and constructively used SEL assessments, like well-developed and constructively used academic assessments, could, should, and increasingly do support educational practice.

In addition, the focus of this article is on SEL assessments designed to be administered to all students, referred to here as *universal* assessment. Just as academic assessments, administered to all students, provide useful information to guide teaching and learning, so too might SEL assessments provide educators with information about all students' social and emotional strengths and needs, and in so doing provide a foundation for programmatic and instructional decision-making to build on strengths and address needs. It is important to note that other social and emotional matters and their assessment are also important, including, for example, the social-emotional climate, teacher–student relationships, and program implementation intensity and quality. The assessment of those and other constructs related to social and emotional

learning are important considerations. It is also significant that SEL assessment for Tiers 2 and 3 and for students with disabilities are important educational tools. Because of the distinctive features and relatively recent advent of universal SEL assessment, however, this article focuses specifically on the universal assessment of student social and emotional competence. Many of the general considerations examined here also apply to other forms of assessment relevant to the field of SEL.

Currently Available SEL Assessments

As I have described elsewhere (McKown, 2017b, 2019), student social and emotional competence assessments encompass several methods, each of which has strengths and limitations. This section reviews the most common methods of SEL assessment, including self-report, rating scales, and direct assessment. Lists of currently available SEL assessments, along with information about their psychometric properties, can be found at CASEL (2019) and RAND (n.d.).

Self-Report

Self-report is often used to assess student social and emotional competence. With this familiar format, children are typically presented with a series of statements about a social and emotional competence and asked to rate on a Likert-type scale how frequently they engage in the behavior or how true the statement is of them. Self-report assessments have several advantages. First, they are comparatively easy to construct and revise. Second, they are easily administered to large numbers of students on widely available survey platforms. Third, self-report questionnaires, more than other forms of SEL assessment, reflect "student voice," providing students an opportunity to provide their views of their SEL strengths and needs. This is particularly important for social-emotional competencies that reflect attitudes and beliefs, such as self-efficacy (the belief that I can accomplish even challenging tasks) or growth mind-sets (the belief that ability is a function of effort).

Self-report questionnaires also have limitations. First, because children can often infer from the item content what a desirable response is, they can be vulnerable to "social desirability response bias" (Crowne & Marlowe, 1960). Second, self-report relies on children's appraisal of their own competencies, which presupposes a level of self-awareness that may not always be present. Third, because of reading and cognitive demands, it is difficult to administer self-report questionnaires to young children, although creative methods for doing so have been developed (Measelle, Ablow, Cowan, & Cowan, 1998).

Likely because of their relative simplicity, self-report questionnaires are widely used to assess student SEL skills. For example, in California, a consortium of school districts applied for a waiver from the U.S. Department of Education to develop and implement an alternative School Quality Improvement System. This consortium, known as the CORE Districts, uses a quality improvement index that includes academic indicators, which make up 60% of the overall index, and social-emotional and climate factors, which make up 40% of the overall index. Social-emotional indicators include self-reported growth mindset, self-efficacy, self-management, and social awareness. CORE Districts use these scores, in combination with other social and emotional indicators, to direct resources, including peer support and focused professional development, to support school improvement (West, Buckley, Krachman, & Bookman, 2018).

Rating Scales

Rating scales are also frequently used to assess student social and emotional competence. With rating scales, an adult, usually the classroom teacher, reads statements describing behavior and rates the frequency with which a child engages in those behaviors. Rating scales were created mainly to assess behavioral problems, although even early in their development, rating scales included scales focused on adaptive or positive behaviors (Achenbach, 1978; Spivack & Levine, 1964). Rating scales have evolved to include assessments that focus on student strengths. Some focus exclusively on student strengths, and some focus specifically on social emotional skills as reflected in the CASEL model (Gresham & Elliott, 2017; LeBuffe, Shapiro, & Robitaille, 2018).

The benefits of rating scales lie in their accessibility and their ability to capture observed behavior across a range of situations and settings. They can be feasibly administered to teachers to rate a large numbers of students, often on computerized survey platforms. Because teachers have daily contact with a large number of same-age peers, they have a strong reference group from which to rate individual student behavior. In addition, to the extent that rating scales ask teachers to rate observable behaviors, they do not require a high level of inference for teachers to assess students.

Although they are straightforward, completing lengthy questionnaires on a large number of students can pose a burden on teachers. To reduce this burden, shorter versions of some rating scales are available, but this also reduces score reliability and the information gleaned from results. In addition, different teachers may rate the very same behavior differently or in other ways act idiosyncratically in their ratings. For example, raters may be vulnerable to "halo" effects by which teachers rate children they

like more favorably than other children with similar behaviors. In addition, teachers many demonstrate leniency or severity biases, in which the teacher rates all children negatively or positively (Merrell, 2009). In addition, because rating scales are best suited to assessing observable behaviors (Merrell, 2009), less visible social and emotional competencies, such as awareness of others' feelings and beliefs, may be more difficult for teachers to observe and therefore accurately rate.

Direct Assessment

Direct assessment is also used to measure student social and emotional competence. With direct assessment, students demonstrate their SEL competencies through solving challenging social and emotional tasks. Some refer to this method as "performance assessment," but that term often refers specifically to the ability to perform a task or demonstrate a competence in a naturalistic context. Direct assessment includes naturalistic performance tasks and other forms of assessment that require children to demonstrate competencies, such as individually administered clinical assessments and group-administered computer-based assessments. A direct assessment that involves naturalistic tasks is the Preschool Self-Regulation Assessment. The Preschool Self-Regulation Assessment consists of a number of tasks designed to measure several dimensions of self-regulation. For one task, for example, a child is asked to wait with an M&M on his or her tongue for as long as possible (Smith-Donald, Raver, Hayes, & Richardson, 2007). Other direct assessments involve the presentation, in one-on-one testing or via computer, of tasks that require children to demonstrate skills. Clinical assessments, such as the NEPSY (Korkman, Kirk, & Kemp, 2007), include subtests assessing children's social awareness, specifically affect recognition and theory of mind skills. The NEPSY is administered one-on-one to children and requires specialized training to administer, score, and interpret.

Computerized direct assessments provide developmentally appropriate tasks that require children to demonstrate their knowledge and skills. These can be game-like tasks that don't appear much like an assessment (DeRosier & Thomas, 2019) or interactive modular assessments with illustrated and narrated item content (McKown, Russo-Ponsaran, Johnson, Russo, & Allen, 2016). For example, SELweb (McKown et al., 2016) includes an emotion recognition module in which children look at faces and indicate what each person is feeling from their facial expressions. To assess social perspective-taking, children listen to illustrated and narrated vignettes and answer questions that require them to infer a story character's intentions.

Direct assessments provide evidence of social emotional competence unmediated by the potentially biased perceptions of a rater or the child herself. Their potential to measure social and emotional skills objectively is therefore a strength. When constructed well, direct assessments can yield highly reliable scores that are valid for a variety of purposes. For direct assessment to be suitable for use in schools, particularly for universal application, they must be simple to administer, suitable for group administration, and offer automated scoring and reporting. The costs of developing and validating such school-appropriate direct assessments is high, and the process is technically complex. In addition, direct assessment may be better suited to assessing the knowledge and mental processes that are involved in social interactions rather than social and emotional skills expressed in behavior. Because of development cost and complexity, few technically sound direct assessments with these characteristics are available to educators.

Other methods of assessment can be used to assess social and emotional competence. For a variety of reasons, they seem unlikely to be adopted widely in education settings. Behavioral observation systems are costly and time-consuming methods of assessing child behavior that are vulnerable to rater bias and may yield information about situation-specific behaviors that do not generalize (Merrell, 2009). Peer nomination techniques are well-established and potentially offer powerful information about peer acceptance and child behavior. However, they require expertise to administer, score, and interpret and are objectionable to many communities (McKown, Gumbiner, & Johnson, 2011). Direct behavior ratings (Christ, Riley-Tillman, & Chafouleas, 2009) have great potential for progress monitoring but are not designed to assess the broad range of social and emotional competencies in the CASEL model. Some school districts use administrative records about absences and in-school suspensions as indicators social and emotional competence. However, these indicators do not measure social and emotional competence itself, but proximal and distal outcomes of those competencies, and so their value for understanding student social and emotional competence may be limited. The reader interested in a more comprehensive catalog of available tools is encouraged consult the CASEL assessment guide (https://measuringsel.casel.org/assessment-guide/) and the RAND assessment finder (https://www.rand.org/education-and-labor/projects/assessments.html).

The growing interest in SEL assessment has led to an increasing pace of assessment development and use, raising several challenges and opportunities. A working hypothesis underlying this article is that the extent to which these challenges are addressed will affect the extent to which the field continues to develop useful assessments

that will be well used and, in so doing, will support teaching and learning. As such, clarifying and addressing each area of challenge and opportunity has the potential to advance the quality of assessments and their use.

CHALLENGES AND OPPORTUNITIES

Challenge and Opportunity 1: Integrating Developers' and Practitioners' Priorities

One challenge and opportunity involves balancing and integrating the differing priorities of assessment developers and assessment users. Assessment developers and assessment users largely focus on the same concerns but place those considerations in different priority order, sometimes resulting in a misfit between the best qualities of SEL assessments and the greatest needs of their users. It is as if assessment developers and users were using the same words but speaking different languages. The challenge here is how to place priorities of assessment developers and assessment users on equal footing so that the SEL assessment endeavor reflects the best combination of rigor and relevance.

To understand the divide, consider a cartoon: In 1976, the cover of *New Yorker* magazine featured an illustration by Saul Steinberg (1976) called "View of the World from 9th Avenue." It featured a bustling, large, and highly detailed Ninth Avenue in the foreground and a somewhat distant and vague Tenth Avenue behind it; far in the background, one can make out New Jersey and other U.S. states and cities, flanked to the south by Mexico and to the north by Canada, with the Pacific Ocean and the distant land masses of China, Japan, and Russia beyond. The illustration is making fun of the tendency of some New Yorkers to see Manhattan as the center of the universe and everything else as beside the point. I bring this up not to insult New Yorkers but because the image illustrates the problem of different perspectives of SEL assessment developers and users.

In many assessment developers' view of the SEL assessment world, rigor—defined as technical quality, score reliability, evidence of validity, strength of norms, and the like—looms large in the foreground, whereas considerations of relevance, defined by the assessment's usability, feasibility, usefulness, relevance to practice, and cost-effectiveness are present but in the background, perhaps where Canada, Mexico, and the Pacific Ocean are on the *New Yorker* cover. That viewpoint can be seen clearly in the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014), which outlines key principles and standards of quality in the design, validation, and use of educational and psychological tests. The Standards provide important guidance to the field and help ensure that assessments meet rigorous quality metrics and are used appropriately. The standards cover a lot of ground and are heavily weighted to describing the desirable technical properties of assessments and the evidence needed to demonstrate those properties. Considerations of relevance, such as how tests can appropriately be used, are included too, but these are not as prominent as considerations of rigor.

The view of SEL assessment from the schoolhouse is quite different: For educators, looming large in the foreground are considerations of relevance, such as whether the assessment can solve an important problem of practice, whether it will help them do their work better, whether it is usable and feasible, and whether the things that are assessed will help support student achievement. Important but less immediate, and therefore in the background, are matters of rigor, as just described. Rigor matters, but only if relevance is high. In other words, the same issues are in the picture for assessment developers and practitioners, but their priority and prominence differ drastically. Of course, rigor and relevance do not exist as a zero-sum proposition—they can be coequal considerations.

There is reason to believe that the developer's view, not the practitioner's, predominates: Educators are starting to adopt SEL assessments, but reviews are mixed. For example, a national survey of principals found that even though most principals (71%) believe SEL competencies can be assessed, (a) only a minority (24%) assess all students' SEL development, (b) most who use SEL assessment (60%) do not find the assessments to be very useful, and (c) most believed their teachers do not know how to use SEL assessment data to inform their practice (DePaoli, Atwell, & Bridgeland, 2017). Similarly, in a Gallup (2018) survey, most educators (87%) reported it is important to measure "nonacademic" skills, but only one in 10 reported that assessments used in their school measured those skills very well. These findings suggest a gap between educators' desire to assess social and emotional competence and their perceptions of what is available to them. Although the survey does not explain where that gap originates, two potential sources seem likely: (a) Educators may be unaware of assessment options that may meet their needs, and (b) available assessments may not yet meet practitioners' relevance needs. A third source is also possible: Available assessments may still not be sufficiently rigorous to meet practitioner needs. Whatever the source of the challenge, SEL assessment has not found its way into routine practice, despite clear practitioner need for these tools.

Reasonable people may disagree with my depiction of the researcher's view or point out that it is overly exaggerated. There is little doubt that important exceptions to this depiction can be found. Rigor and relevance: Neither

is better; both are important. And so as the field of SEL expands and greater attention is paid to the issue of how in the world to assess the vast range of skills called SEL, assessment developers and users will benefit from striving for a consistent balance between rigor and relevance, with roles and responsibilities falling to assessment developers and users alike.

The Opportunity

The question for the field, then, is how assessment development efforts can proceed in a way that will combine the best of psychometric rigor with the practical relevance needed to make SEL assessment viable tools that support teaching and learning. Several strategies will help hew rigor to relevance. First, involving the intended end user in the assessment development endeavor from the design phase is one practical step assessment developers can consider to keep relevance in the foreground as they design or improve SEL assessments. At each step in the assessment design and field-testing cycle, structured user input can help shape the elements of the assessment. For example, before a prototype is developed, when assessment developers have defined what they want to measure (preferably in response to, or at least with input from, the field) and the assessment's intended use, the users can provide input about how important, relevant, and useful the assessment concept would be to practitioners and what would make it more important, relevant, and useful. They can also provide invaluable insights into what the barriers to the intended use might be.

Next, when an assessment is prototyped—including when pilot items, response options, delivery format, and user interface are developed—users can provide input on the usability and feasibility of those features. At this point, the assessment developer might also prototype the score reporting format based on the intended interpretation and use and seek input from the intended end user. This input can be collected, conceivably in iterative cycles, and used to revise the prototype until the assessment design has met a standard of usability and feasibility that gives the completed version a reasonable chance of being useful to the intended end user.

Next, when the assessment is developed and brought to field testing, in addition to gathering data on the technical properties of the SEL assessment, the assessment developer can collect, in parallel, user input. User input surveys might be used to obtain feedback on usability issues like ease of administration, duration of assessment data collection, as well as asking for information about barriers to use during the field trial, and suggestions for improvements. Armed with data on the assessment itself, and factors influencing its usability, the assessment developer can proceed with revisions that improve the assessment's technical qualities (rigor) and its usability and feasibility in real-world settings (relevance).

Finally, until schools of education include more training in SEL programing and assessment, when SEL assessments are rolled out to the field, developers and users alike should anticipate that professional learning opportunities will be needed for those assessments to realize their most constructive potential. Depending on the assessment goals, context, and intended uses, this might include teacher training in what SEL is, why it matters, the goals of assessment, the properties of the chosen SEL assessment, the relationship between assessment and instruction, and training on interpreting and using SEL assessment scores for decision-making.

Challenge and Opportunity 2: Interpretation and Use of SEL Assessment Data

A particular concern in the school-based use of student SEL assessment is the suitability of the assessment for the inferences and decisions educators plan to make based on assessment scores. When it comes to SEL assessment, to a large degree, psychometric merit should be judged in the same way any educational or psychological assessment would be judged (AERA, APA, NCME, 2014). However, one psychometric consideration, in my view, is particularly important for the field of SEL assessment to contemplate as it moves forward: When considering an SEL assessment for an intended set of interpretations and uses, what evidence supports the use of that assessment for those interpretations and uses? This is important because highly consequential decisions, such as what to teach, how to teach students, student educational placement, and policy decisions, to name a few, should be based on information that is appropriate for making such decisions. If the evidence does not support the intended use, faulty decisions can result.

Here are some specific examples of the kinds of decisions that SEL assessments can be used to make. A teacher might administer an SEL assessment before starting a series of SEL lessons. She might use what she learns about student strengths and needs to decide what lessons to emphasize and what skills to reinforce. This reflects the formative use of assessment. A teacher or administrator might assess social and emotional competence after a period of SEL instruction to measure progress. That is summative assessment. An investigator might use SEL assessment as an outcome measure in a field trial of a new intervention—that is program evaluation. A district might use social and emotional competence assessment as part of an index of school quality to determine where to direct resources for school improvement—that is a form of accountability. Another district might use change in measured social and emotional competence across a year

to create a "value added" measure of teacher quality linked to student social and emotional competence—that is a higher stakes form of accountability.

The appropriateness of a given assessment for each of these uses can be judged based on evidence of the assessment's properties. Only when the evidence supports a given use should an assessment be used in that way. This is true of academic assessment too. However, there are reasons to take particular care in considering the manner by which SEL assessment data will be interpreted and used with particular care.

Why Interpretation and Use Is Particularly Important for SEL Assessment

Three facets of the policy and practice context makes interpretation and use particularly important considerations for SEL assessment. First, whereas teachers and other professionals in the education system are, generally speaking, academic subject matter experts who are therefore in a good position to make valid inferences and decisions based on achievement data, they may not have such content expertise in SEL. SEL is not commonly a part of teacher induction programs (Schonert-Reichl, Hanson-Peterson, & Hymel, 2015), nor is it typically part of teacher in-service training. In a 2014 national survey of teachers (Bridgeland, Bruce, & Hariharan, 2014), 23% of teachers reported receiving any in-service professional development, including professional development focused on SEL. This situation increases the risk that they might interpret and use SEL assessment data in a way that is not supported by the evidence.

Second, the SEL policy context creates uncertainty. Consider a contrast: Educational policies and well-established local practices often make very explicit when achievement tests are to be used, for what purpose, and what inferences or decisions are to be made based on the score results. For example, mandatory statewide achievement testing dictates what assessment will be administered, what its content will contain, and how the assessment data will be used. Whether educators like these assessment constraints or not, they provide educators quite a bit of clarity about what is to be done. In contrast, SEL has only recently appeared in any state policy, and no policy requirements dictate whether educators must assess SEL, what they must assess, what assessment they must use, why they must assess it, or how the assessment data will be interpreted and used for decision-making. As a result, there is a greater burden on local school districts and individuals within the districts who plan to define these parameters. In this context especially, without strong guidance and requirements, there is greater room for untoward interpretation and use of SEL assessment data.

Third, the field of SEL assessment development for broad educational use is relatively young. Contrast this with the more established field of educational testing, which has a very long history, track record, and standards and practices to guide judgments about quality and use. Because the field of SEL assessment is young, assessment forms and content may be unfamiliar, and the accumulated evidence of the properties of those assessments is limited. As a result, it may be more difficult to ascertain what constitutes a good SEL assessment and what inferences are supported by the data.

Responsibilities of Assessment Developers and Users

Because of these factors, a particularly acute challenge facing educators in any district contemplating using SEL assessment concerns what Messick (1995) called "consequential validity." Specifically, educators who want to use SEL assessment as productively as possible with the fewest negative unintended consequences need to be confident that the intended interpretation and use of SEL assessment data are justified by the evidence. To increase the odds that an SEL assessment will be used to make inferences and decisions that are justified by the data requires the participation of assessment developers and users alike.

If Messick (1995) successfully argued that the consequences of assessment interpretation and use should be a validity consideration, Kane (2013) described how assessment developers might go about developing evidence to evaluate consequential validity. Specifically, Kane argued that assessment developers, from design through field testing, are well advised to develop what he called an "interpretation and use argument," or IUA. The IUA makes explicit the intended inferences that assessment users ought to be able to make from scores from a given assessment. The assessment developer can use the IUA to determine what types of evidence are needed to determine how well suited the assessment is to its intended interpretations and uses. The responsibility of assessment developers, then, is to articulate their IUA and to be transparent about what evidence supports the IUA, what evidence does not support the IUA, and what parts of the IUA have insufficient evidence to judge.

This is something akin to a pharmaceutical manufacturer being clear about the indications for a particular medicine. The differences in the professional practices between medicine and education are instructive. In medicine, physicians guide the appropriate use of medicines through prescription privileges. In that capacity, therefore, they serve as an intermediary with the expertise to ensure that medicines intended for a particular condition are used for that condition and (in general) not for others unless

the benefits of such use would eclipse the risks. There is no analogous intermediary in the world of SEL assessment. Educators purchase and use SEL assessment directly from the assessment developer, and therefore are, to extend the analogy perhaps to the breaking point, both the doctor and the patient. As a result, the user has an obligation to read the metaphorical label—to understand what interpretations and uses of a particular SEL assessment are supported by the evidence. Because SEL assessment and its use is not dictated by policy, this obligation is even greater when evaluating SEL assessment than it is when evaluating academic assessment.

Just as a self-prescribing patient runs risks that will not accrue to the patient under a doctor's care, the SEL assessment user who is determining the fit between an SEL assessment and the anticipated interpretations and uses runs risks. The biggest one is using the assessment data to make inferences and decisions that are not justified by the evidence. Examples include diagnosing a child based on an assessment that is not valid for making diagnostic decisions, evaluating teacher performance based on SEL assessment scores over time when there is no evidence to support that use, and labeling a child as "at risk" of an emotional and behavioral disorder when there is no evidence to support that use of an assessment.

To be sure, educational diagnosis, teacher evaluation, and screening are all legitimate and important educational endeavors. In theory, the right SEL assessment could be used to pursue those endeavors so long as the evidence supported that use, although the appropriateness of doing so with SEL assessment even if the assessment is technically up to the job is debatable. The point is that assessment developers have an obligation to provide evidence supporting whatever the intended interpretation and use. And assessment users have an obligation to understand what interpretations and uses are justified by the evidence (APA, AERA, NCME, 2014).

Opportunity 1: Goal Clarity

There are several actions that those contemplating SEL assessment might consider to increase the odds that the SEL assessment they use will be up to the task of their intended interpretations and uses. First, users would do well to identify, before selecting an SEL assessment tool, what their assessment goals are and are not. What do they want to understand about their students? What kinds of decisions do they intend to make from what they learn? Conversely, what kinds of decisions and interpretations are off limits for the SEL assessment project? With those goals in mind, and in consultation with district personnel or an outside consultant with the expertise to evaluate the evidence supporting the interpretations and uses of SEL assessments, district personnel can skillfully evaluate the suitability of the assessment alternatives to achieving their aims.

Judging Evidence

This raises the question of what standards to use to weigh evidence of an assessment's appropriateness for a given interpretation and use. There is no simple answer to this question, and the reader can find excellent treatments of the issues in Messick (1995) and Kane (2013). Two useful principles should guide those judging the merits of an SEL assessment for its intended purpose. First, the user should clarify what inferences and decisions they are and are not going to make with the SEL assessment data. This will help to narrow the search for evidence of the appropriateness of the assessment for the intended use. The evaluative questions, forms of evidence, and metrics for judging quality will then flow from the intended use. Table 1 provides examples of six common intended uses, the evaluative questions users might contemplate, sources of evidence to answer those questions, and general standards by which to judge the sources of evidence. Note that this is for illustrative purposes and that there are doubtless different ways of categorizing intended uses and identifying kinds of evidence and metrics to judge the appropriateness of an assessment for those uses. The key message is that there are many ways that student SEL assessments might be used and users should be clear about their intended use and that the evidence supports their chose assessment for this intended use.

A second important general principle involves the stakes of the inferences and decision to be made from the assessment data. The principle is this: The higher the stakes of the decision to be made based on the assessment data, the higher the standard of excellence that the assessment should meet. Arguably, all educational decisions are high stakes. Nevertheless, some appear to be higher stakes than others, judging by their consequences. Making an informal observation of a student's behavior and adjusting a behavior management strategy for a single instructional period based on that observation is fairly low stakes. The period is limited, and the consequences of the decision do not affect student grades, teacher performance review, student placement, teacher salary, and the like. Informal assessment with minimal evidence of psychometric merit is, in this context, appropriate. In fact, teachers constantly evaluate students informally to adjust their teaching.

On the other end of the spectrum are decisions that have long-lasting effects that may strongly affect student outcomes. For example, diagnostic decisions about special education placement are high-stakes decisions. Data used to make decisions about teacher pay and performance and school funding are high-stakes decisions. Publicly available data that parents use to decide what community to

TABLE 1
Sample Intended Assessment Uses and Criteria for Evaluating Assessments for Those Uses

Intended Use	*Evaluative Question*	*Evidence*	*Metrics*
Infer how student performance compares to the general population.	How well does the normative sample reflect the reference group I care about? How long ago were the normative data collected?	Characteristics of the normative sample and timeframe of normative data collection.	In general, the higher stakes the decision, the better the norming sample ought to be.
Understand student strengths and needs to guide instruction	Does the assessment measure what the instructional program teaches? Are scores useful in guiding instruction? Is performance on the assessment is associated with the skills taught?	Relationship between measured skills and the content of the instructional program's scope and sequence. Score reliability and correlation with other measures of the skills taught.	Assessment content. Internal consistency and temporal stability for score reliability. Correlation with other variables.
Evaluate student progress over time	Is the assessment sensitive to change in student skill level over time?	Evidence that scores increase with age as expected. Evidence of consistent performance on repeated assessment.	Improvements in performance by age with cross-sectional or longitudinal data. Test-retest reliability.
School and teacher continuous improvement	Does performance on the assessment reflect the skills of the people and the quality of their practices?	Associations between known measures of school quality and student skill in the context of longitudinal, quasi-experimental, or experimental designs.	Magnitude of correlation between school quality and performance on the assessment. Data suggesting that continuous improvement lead to improved scores.
Evaluating program impact	Is the assessment sensitive to the impact of effective programs? Is the assessment designed to measure what the program is designed to teach?	See "understand student strengths and needs … " Evidence that the performance of students exposed to high-quality instruction improved more than a control group.	See "understand student strengths and needs to guide instruction." See field trials that included the assessment as an outcome measure.
High-stakes school or teacher accountability	Do schools and teachers produce a measurable "value-add" to student social and emotional growth, such that they can be held accountable for that growth?	Teachers and schools account for variability in student growth. Growth can be attributed to practices. Practices are in educators' control	Variance components analysis over time. Association between teacher practice and between-class and school variability.

live in, to the extent that they concentrate the tax base in "better" districts, are high-stakes uses. In these cases, the consequences of inaccurate data are particularly high, involving the kind of instruction and the restrictiveness of the environment in which students will be educated, teacher livelihood and professional reputation, and the tenor of communities potentially facing school closures or, in the other direction, an influx of tax revenue and prosperity. It seems clear that for these inferences, the quality of the evidence supporting those inferences ought to be very high.

Opportunity 2: Other Actions to Ensure Appropriate Use

Ensuring that interpretation and use of SEL assessment data are supported by evidence will require several conditions. First, the field will benefit from specific guidance about the standards of evidence to which assessments should be held in support of defined intended uses. Ongoing efforts by organizations such as the Buros Center on Testing (https://buros.org/) aim to provide just that kind of guidance. Second, assessment developers should be clear about the intended interpretations and uses of their assessments and should present evidence that supports those uses and, where evidence does not, should provide guidance about the limitations of their assessment. Third, assessment users should understand the supported interpretations and uses of an SEL assessment and plan to use the assessment for those uses only. Conversely, if a user has an intended interpretation and use in mind, as they review assessment options, they should actively seek and request evidence supporting those interpretations and uses for options they are considering.

Challenge and Opportunity 3: Creating Conditions for High-Impact SEL Assessment

A third challenge and opportunity concerns what happens after SEL assessment data are collected. Data can be useful only if users have an opportunity to review, discuss, and interpret their meaning to decide what to do based on what they learned. Although that premise may seem obvious, it is much less obvious that school districts consistently engage in systematic data review and that those that

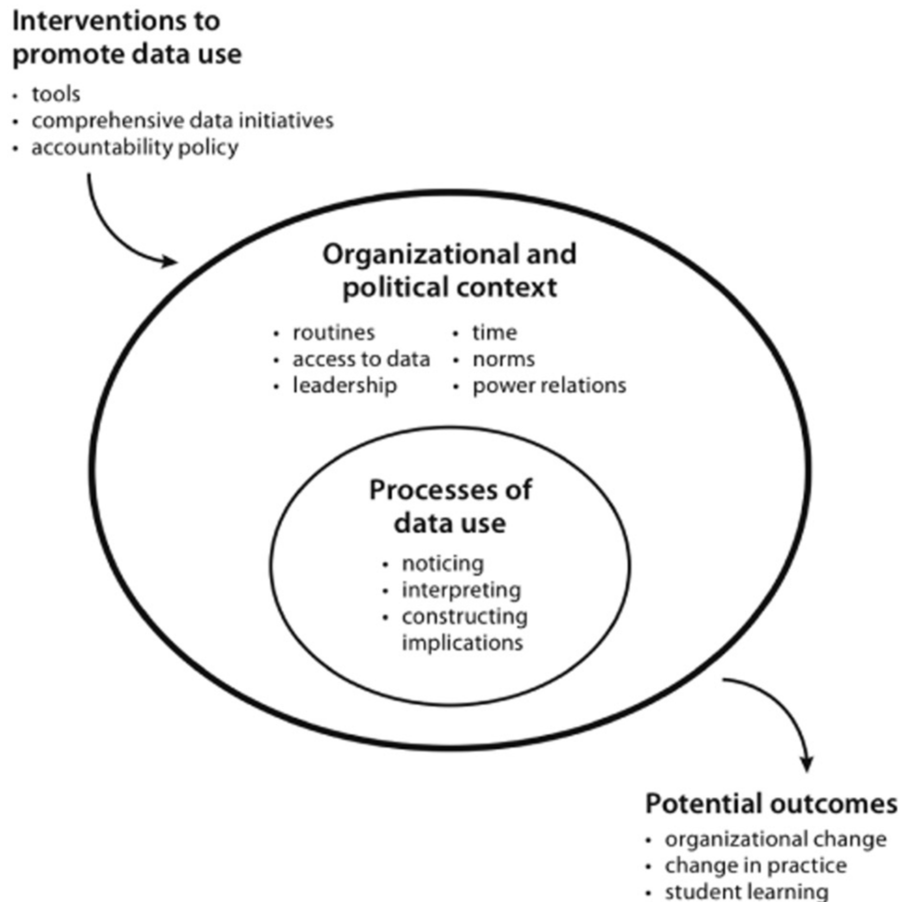

FIGURE 1 Framework for data use. Source: Coburn & Turner, 2011. (© 2011. Taylor & Francis. All Rights Reserved. Image reproduced with permission.)

do review data in a way that is likely to lead to insight and positive action.

Coburn and Turner (2011) summarized research on data use in schools and proposed a model of the factors that influence how data are used, by whom, and to what ends. This model recognizes that schools are complex systems and that different constituents at different levels of a district—including superintendent, cabinet, principal, and teacher—each operate with different goals and incentives. As a result, many complex forces influence how any assessment, including SEL assessment, is selected, reviewed, and used for decision-making.

Coburn and Turner's (2011) framework, depicted in Figure 1, describes concentric contexts that influence the outcome of SEL assessment. Working backward from the outcome, they envision that data have the potential to influence practices and thereby student outcomes. Changes in practice and student outcomes are, most immediately, the result of data use processes engaged in by educators. These "processes of data use," described next, involve practitioner review of and reflection upon the assessment data. Data use processes, in turn, are influenced by dimensions of the organizational context that include routines, access to data, time, norms, leadership, and power relations. Using this framework as a starting point, I describe considerations that are likely to influence the success and impact of SEL assessment initiatives. To Coburn and Turner's framework, I add SEL assessment goals as an important context that will affect data use and impact.

Goals

As described in Table 1, social and emotional assessment, like achievement testing, can be undertaken in pursuit of a variety of goals (McKown, 2019). Because the goals of SEL assessment are in general not as clearly spelled out by policy requirement or tradition as the goals of academic assessment, the onus is therefore on educators planning to assess SEL skills to be clear about what goals they intend to pursue. Goal clarity means clarity about how assessment will be used and how it will *not* be used. As I have written elsewhere (McKown, 2017a, 2017b, 2019), assessment goal clarity, or lack thereof, can be highly consequential. Goal clarity facilitates communication among education professionals, parents, and the

community. For example, when a parent expresses concerns about SEL assessment, if goals are clear and widely agreed upon, it is easy to communicate to the parents why that assessment is being pursued and to anticipate and address concerns. Goal clarity also allows educators to anticipate how assessment will benefit them. For example, in a district that is using SEL assessment for formative purposes and not to evaluate teacher performance, teachers can anticipate that they will obtain useful information about their students that they can use to modify instruction. At the same time, they do not have to be concerned about the impact of assessment results on job security and compensation. In addition, defining SEL assessment goals clearly will shape how data are used (see McKown, 2019).

Defining the User

The first consideration in data use is, who will have access to SEL assessment data? Who the data "users" are depends heavily on the assessment goal. If the goal is formative assessment to influence teacher practices, then clearly teachers are high-priority users. Others charged with supporting teachers, including principals and others, may also be part of the user group. On the other hand, if the goal is program evaluation, the main users will be those charged with making program investments and program development. That will often be members of the district cabinet who will use the data to determine whether their program investments are yielding the desired benefits, though others in the professional community will no doubt be interested in the data as well. The important issue here is that effective data use practices will include clarity about who the users of the data are. That will largely be guided by assessment goals.

Data Use Processes

Defining goals and users set a context for data use, defining who will use the data and toward what ends. But those steps do not themselves constitute data use. Data use happens when users have time to review, reflect upon, and interpret assessment data and commit to actions they will take based on what they learned. Coburn and Turner (2011) called this the "process of data use," and it includes noticing, interpreting, and constructing implications based on the data. Noticing data involves ascertaining the facts from the data—such as what scores children achieved and what skills those scores reflect. Interpreting refers to making inferences about the meaning of the scores. And constructing implications refers to developing ideas of what to do based on what users notice and what their interpretations of the scores are. The process of data use requires precious resources—in particular, time, knowledge, and norms for facilitating discussion.

The easiest, and likely most feasible, resource for noticing, interpreting, and constructing implications for SEL assessment data are routines that already exist and to which SEL assessment data may be added. Imagine, for example, a district that regularly convenes grade-level team meetings in which reading and math assessment data are reviewed and used to make instructional decisions. In that school, assuming this is consistent with their SEL assessment goals, it would be relatively straightforward to add SEL assessment data so that the team meetings now involve reviewing student reading, math, and SEL status. Given the many demands on educators' time and energy, creating new meetings, and reviewing SEL assessment data apart from achievement data, runs the risk of inadvertently communicating that SEL is something distinct and separable from academic endeavors, and something that poses additional burdens of time and effort with no clear payoff.

In districts without clear and routinized data review processes, the challenge is somewhat greater. In this case, time, the most precious of resources, needs to be found to engage in data review. In addition, norms for data review discussions need to be established so that the data are interpreted within the limits of evidence supporting their interpretation and use argument. Finding time more likely means commandeering time used for one activity to make room for SEL assessment data review. What time is committed for data review by the intended users requires the sanction of leaders at building and district levels who prioritize the use of staff time. Similarly, establishing norms and practices for reviewing and interpreting SEL assessment data requires leadership and, likely, professional development. And so finding time and establishing routines for SEL assessment data use, particularly when they need to be developed from the ground up, requires leadership, an important context of use.

Linking Data to Action

An important element of the data review process is deciding what actions to take based on insights gleaned from the assessment data. Educators are subject matter experts in academic subjects and are therefore well equipped to make decisions about how to use achievement data to modify instruction. In contrast, they may be less familiar with SEL and therefore less equipped to use SEL assessment data to make decisions about how to modify their practice for the benefit of students.

Many schools that opt to assess student SEL will have at least begun the process of implementing some form of SEL instruction. This may be, for example, in the form of an SEL curriculum for use in general education

(Weissberg, Goren, Domitrovich, & Dusenbury, 2012). Assuming that the SEL assessment is designed to measure skills that are the subject of the curriculum, there is a natural connection then between what is assessed and what is taught. Educators can use what they learn from SEL assessment to decide how to use the SEL program scope and sequence, for example, emphasizing units that teach skills that appear, from the assessment data, to reflect particular areas of need.

Few SEL programs include student SEL assessments and few student SEL assessments include curricular or instructional materials. Therefore, it may take some work, even when the SEL assessment and programs appear to cover very similar skill areas, to identify the correspondence between the units and lessons in the SEL program and the scores obtained from the SEL assessment. This work will require leadership to make time and professional learning available so that users can engage in a meaningful process of reviewing, reflecting on, and making decisions based on what they learned from the assessment data.

The Data Use Context

Data use clearly cannot happen in a vacuum. Even time and norms and processes for interpretation—the most basic requirements for data to influence educators' thinking and practice—cannot be taken for granted. Creating data use routines requires as much leadership as deciding to assess student SEL skills. The most successful SEL assessment initiatives will have the full support of the district leadership and will be provisioned with sufficient time, personnel, and professional development to ensure that going to the trouble of collecting student SEL assessment data is likely to influence teacher practices and student outcomes.

The Opportunity

Schools and districts will naturally vary in the extent to which the conditions of high-quality data use are in place. As a result, when embarking on SEL assessment initiatives, educators would be well advised to begin with a self-assessment to determine the extent to which leadership supports SEL assessment, assessment goals are clear, users are defined, routines and practices are in place for data review, and resources and supports are available to take positive action based on assessment findings. Furthermore, because SEL is outside of the content expertise of many educators, professional learning opportunities on the meaning, interpretation, and use of SEL assessment data may be necessary to ensure effective data use. This self-assessment can serve then as the basis for mobilizing resources necessary to ensure that SEL assessment data will be used and that their use will be effective.

Challenge and Opportunity 4: Coordinating Standards, Programs, Assessment, and Professional Learning

One of the more significant challenges facing the field is the lack of coordination between standards that indicate what social and emotional competencies students should know and be able to demonstrate, SEL assessment, SEL instructional programs, and professional learning. It is likely that if standards, assessment, programs, and professional learning were coordinated, this would yield several benefits. First, SEL standards would communicate clear expectations about student social and emotional competencies are most important. This would help guide educators' instructional efforts to teach the same competencies. Second, SEL assessment developers would have clear guidance about what content areas their assessments should be designed to measure. Assessments designed to measure the competencies described in standards are in the best position to support educators as they work to help students meet those standards. Similarly, SEL instructional programs could be designed to support student mastery of the competencies designed in the standards. Fourth, assessments and programs could work in concert to support student SEL, because the assessments would be designed or refined to measure the same competencies the programs are designed to teach. Fifth, professional learning—from preservice through in-service—would develop educators' social and emotional content expertise so that they were better positioned to use and interpret SEL assessment data and effectively implement instructional strategies designed to address student social and emotional learning needs. These anticipated benefits of coordinated standards, assessments, programs, and professional learning reflect hypotheses; a corollary hypothesis is that without such coordination, the potential of SEL and its assessment, will be limited.

In terms of standards, although all states have standards for early childhood social and emotional learning, most states have not developed guidelines or standards indicating what social and emotional skills children should know and be able to demonstrate from kindergarten through high school (Dusenbury et al., 2018). Arguably the most important context for the coordinated and effective use of SEL assessment, and SEL programs and practices, is therefore lacking. However, a growing number of states have adopted or are in the process of considering social and emotional standards that clarify what children at different grade levels should know and be able to do. As of September 2018, 14 states had specified the social

and emotional competencies that are important to their students in preschool through 12th grade (Dusenbury et al., 2018). In those states, then, district decision-makers can use standards to review and select the SEL assessments and program they feel will be most able to support student acquisition of the social and emotional competencies described in their state's standards.

Educators in the rest of the country face a lack of clear policy guidance on what specific and high-priority SEL skills should be assessed and addressed. However, to the extent that states are adopting clear and actionable standards, educators in states without SEL standards may look across state lines for guidance. However, in those states, the onus for being clear on what to teach, and how to teach and assess it falls generally on superintendent-level district leaders. In most states, then, educators wanting to assess and address children's social-emotional learning must spend extra effort to decide what competencies are most important, what SEL assessments are best suited to assessing those competencies, and what programs are best suited to teaching them.

Without broadly agreed-upon expectations of the social and emotional competencies children should know and be able to demonstrate, assessment developers and program developers cannot use this policy signal to develop assessments and programs that support meeting those expectations. Furthermore, coordinating SEL assessment and instruction is difficult at best: Most widely available SEL programs do not currently include an SEL assessment and most SEL assessments do not include SEL program guidance. As a result, as districts work to identify a preferred SEL program or instructional approach, if they want to use SEL assessment to support their work, they must then identify an SEL assessment that measures the SEL skills that matter most in a way that supports instruction. Whether or not they are in states with SEL standards, because SEL assessment and SEL programs are generally less familiar to educators than academic assessment and programs, there may be a greater need for professional learning to support these efforts.

As we described previously, in the section on conditions of data use, very few teacher induction programs focus on SEL (Schonert-Reichl et al., 2015) and ongoing professional development in SEL is also limited. This means that on their first day on the job, educators lack the expertise to effectively assess and address student social and emotional learning. This means that educators who want to engage in SEL assessment and to teach SEL have to acquire the competence to use and interpret assessment effectively, to teach SEL competencies and, ideally, to use assessment data to guide instruction.

The Opportunity

Working in concert, policy, SEL assessment, SEL programs, and professional learning provide a context in which SEL assessment has a clear and compelling purpose, to support instruction toward standards. What might it look like for policy, SEL assessment, SEL programs, and professional learning to be coordinated? First, through standards and guidelines, states would express clear expectations about the social and emotional competencies students at different grades should know and be able to demonstrate. This would result in investment by developers to create or refine assessments and programs that reflect those standards. Assessments designed to measure the same standards-based competencies educators intend to teach could strongly support effective instruction. Second, schools of education would include courses on social and emotional development, assessment, and instruction as a regular part of the teacher induction curriculum. In this way, from their first day on, educators would have developed subject matter expertise in SEL they can use to assess students, interpret assessment findings, and use what they learn to engage in effective SEL instruction. Fourth, ongoing professional learning would support educators to build on and update their expertise in assessing and addressing student social and emotional competencies. In this way, standards, assessment, and classroom practices would be coordinated and focused on supporting the same student outcomes, and educators would have the expertise to put strong SEL assessment and instructional programs to good use.

What's this all got to do with SEL assessment? A key underlying premise of this article is that SEL assessment will be most effective at supporting high-quality teaching and learning when it is undertaken in service to a clearly defined and meaningful educational goal. When standards, programs, assessment, and professional learning experiences are systematically coordinated, this provides just such a context, one in which SEL assessment is best positioned to support teaching and learning.

Challenge and Opportunity 5: Unwavering Assessments in Varied Cultural Contexts

A final challenge and opportunity in the assessment of student SEL skills concerns culture. A major contributor to this challenge is a tension between traditional assessment development processes that strive for standardization and the cultural variation in the populations in which those assessments are used. On one hand, assessments are designed to measure skills consistently and with as little unnecessary variation (error) as possible. This requires a process of standardization that often results in item types, formats, and content that are quite specific and vary little

from one item to the next. On the other hand, a multicultural society contains a high degree of heterogeneity in beliefs, attitudes, and lived experiences. The contrast between unwavering assessments and cultural heterogeneity poses important challenges and creates unique opportunities.

Specific challenges include questions about (a) concerns about test bias, (b) the frame of reference from which assessment content originates, and (c) concern about the reproduction of test score inequalities and stigma in a new domain. These challenges are arguably more acute in SEL assessment than in academic assessment—whereas most would agree, for example, that reading and math skills are important for all, different people may argue that different SEL competencies are most important, and culture may be one of the factors shaping what skills are valued by whom.

Test Bias

A concern of in any assessment involves systematic bias wherein either items or test scores. With differential item functioning, the question here is, Do children from different groups and who have the same skill level tend to get the same score on each item? If not, the item displays differential item functioning and may need to be revised or thrown out to ensure the overall test is unbiased (Millsap & Everson, 1993). The second is test score equivalence, typically assessed using confirmatory factor analysis methods. There are three particularly important kinds of test score equivalence (Millsap, 2011). The first is *configural* invariance, which means that the tests measure the same things in all groups. Next is *metric* invariance, which means that a 1-point score difference means the same thing for children from different groups. Finally there is *scalar* invariance, which means that children with the same skill level achieve the same score on the assessment. Ideally, an assessment will demonstrate configural, metric, and scalar invariance. To the extent that an SEL assessment is submitted to rigorous tests of differential item and test functioning and appears to function similarly for different groups, it can be described as bias free. To the extent that it does not, limitations to interpretation and use should be disclosed to users.

Frame of Reference

Testing for differential item and test functioning is an incomplete answer to the question of cultural appropriateness. This is because in a culturally diverse society, views of what SEL competence is, how it is enacted, and how it should be measured may differ for people from different cultural communities. Measurement equivalence can address only what the assessment affirmatively measures. But what is *not* in the assessment, and particularly whether something important has been left out, is an important consideration too. What is not measured cannot be empirically tested through even the cleverest tests of measurement equivalence. This matters because it is possible that important social and emotional competencies that are critical to particular cultural communities may not be measured by a given SEL student assessment. To the extent that important SEL skills are *not* represented in a student SEL skill assessment, even if that assessment meets rigorous standards of measurement equivalence, the omission of culturally salient skills may undermine its universal relevance.

Jagers, Rivas-Drake, and Borowski (2018; this issue) offered a helpful cultural analysis in which they recast SEL in terms of the competencies required to promote a more just and equitable world. Specifically, they framed CASEL's definition of SEL to prioritize competencies that might be expected to advance social justice and equity. For example, they suggested that self-awareness include racial-ethnic identity, that social awareness include recognition of race-related social communications, that responsible decision-making include working toward equity in community and classroom, and that relationship skills include cultural fluency and code-switching, or being able to cross between cultural contexts with different rules of behavior. This framing of SEL competencies has profound implications for what and how assessment and programs might be developed.

An important question, then, is how a universal and "color-blind" approach to defining SEL might have limited the field so far. On one hand, there is good evidence that the SEL skills reflected in CASEL's model, in widely used SEL programs, and in state standards are relevant to a broad range of cultural groups: The most comprehensive meta-analysis of SEL programs (Durlak et al., 2011) found that SEL programs, when implemented well, lead to improved student outcomes. As part of their analyses, Durlak and colleagues (2011) looked for, but did not find, any evidence that the benefit of these programs was different for children from different ethnic or socioeconomic groups. This suggests that the SEL skills that are the focus of many SEL programs and practices are important for all children and that assessing and addressing those skills will benefit all children. On the other hand, other SEL competencies, or cultural variants of these competencies, such as those described by Jagers et al. (2018), may be important to assess and address. The question of what SEL skills should be assessed and addressed is an important consideration as the field continues to evolve.

Highlighting Inequalities and Creating Stigma

Some are concerned that SEL assessment might be used in ways that stigmatize children (Edge Research, 2018).

Looming large is the concern that by measuring student SEL skill, educators will document new kinds of racial or gender gaps and associated deficits that unproductively echo a cultural narrative about inequalities. In addition to causing emotional pain, information about mean differences on SEL assessments between members of different groups can cause problems by, for example, subtly shaping what teachers expect from and how they treat their students, producing stereotype threat and self-fulfilling prophecies (McKown & Weinstein, 2003, 2008).

The Opportunity

There is no complete answer to these concerns. However, most SEL assessments will not have been developed for the purposes of highlighting group differences. In addition, most SEL assessment users' primary intended use will not be to highlight group differences. Assessment developers and users, then, have a responsibility for preventing the repetition of unproductive cultural dynamics that highlight, reinforce, and reproduce inequality. For assessment developers, the intended uses argument should be explicit, then, about the appropriateness of using an SEL assessment for describing group differences, and the evidence that supports that use. Assessment users, if they intend to use SEL assessment data to describe group differences, should be prepared to justify the reason for doing so, and that justification should include a solving a specific problem of practice, examining group differences.

A related concern regarding culture and SEL assessment is the worry that SEL assessment data will shine a light on individual skills when that light might be more productively shined on characteristics of the setting. In a district with differential rates of disciplinary referrals for students from different ethnic groups, for example, some might be concerned that school leaders will use SEL assessment data to explain the discipline gap as arising from student SEL skill deficits while discounting the role of adult practices in creating the discipline gap.

Any tool, including SEL assessment, can be assimilated into in ill-conceived or harmful purpose, such as inappropriately attributing inequity to the skill deficits of a group. There is no characteristic of the assessment itself that can prevent this from happening. However, if educators place an equal focus on assessing adult practices, the conditions of learning, and student competencies, and use assessment data about all three to improve practice and student outcomes, this may prevent an overly myopic focus on student competencies at the expense of reflecting on the adult practices that are intended to foster those competencies.

CONCLUSION

Like any major initiative in a multicultural society, the prospect of universally assessing student social and emotional competencies has the potential to make things better, or not. The current state of the field suggests that there will be continued growth in efforts to develop and deploy usable, feasible, and technically sound methods to assess student SEL skill. The progress of this work will be influenced by the factors described in this article including the extent to which (a) SEL assessments balance psychometric rigor with practical relevance; (b) assessment developers and users take steps to ensure that the inferences and decisions that an SEL assessment will be used to make are supported by the evidence; (c) educational leaders support and create conditions by which educators can productively use SEL assessment data for decision-making; (d) standards, assessment, programs, and professional learning are coordinated to support one another and student outcomes; and (e) assessments are appropriate for the varied cultural contexts in which SEL assessments are used.

There are undoubtedly other factors that will influence the state and direction of the field of SEL assessment. Regardless, a reasonable hypothesis is that the more these challenges are constructively, systematically, and effectively addressed, the more widespread and constructive will be the use of SEL student assessment for improving teaching and learning. The education research community will play a key role in testing and refining these propositions about the contexts, potential benefits, and risks associated with student social and emotional competence assessment.

FUNDING

The research reported here was supported by Institute of Education Sciences through Grant R305A160053 to Rush University Medical Center. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID

Clark McKown http://orcid.org/0000-0001-9694-1179

REFERENCES

Achenbach, T. M. (1978). Psychopathology of childhood: Research problems and issues. *Journal of Consulting and Clinical Psychology*, *46*(4), 759–776. doi:10.1037/0022-006X.46.4.759

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint

Committee on Standards for Educational, & Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Aspen Institute National Commission on Social, Emotional, and Academic Development. (2019). *From a nation at risk to a nation at hope.* Washington, DC: The Aspen Institute. Retrieved from https://www.aspeninstitute.org/programs/national-commission-on-social-emotional-and-academic-development/

Bridgeland, J., Bruce, M., & Hariharan, A. (2014). *The missing piece: A national teacher survey on how social and emotional learning can empower children and transform schools.* Washington, DC: Civic Enterprises. Retrieved from https://www.casel.org/wp-content/uploads/2016/01/the-missing-piece.pdf

Christ, T. J., Riley-Tillman, T., & Chafouleas, S. M. (2009). Foundation for the development and use of direct behavior rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, *34*(4), 201–213. doi:10.1177/1534508409340390

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173–206. doi:10.1080/15366367.2011.626729

Collaborative for Academic Social and Emotional Learning. (n.d.-a). Overview of SEL. Retrieved from https://casel.org/overview-sel/

Collaborative for Academic Social and Emotional Learning. (n.d.-b). What Is SEL? Retrieved from https://casel.org/what-is-sel/

Collaborative for Academic Social and Emotional Learning. (2019). CASEL assessment guide. Retrieved from https://measuringsel.casel.org/access-assessment-guide/

Corcoran, R. P., Cheung, A. C. K., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, *25*, 56–72. doi:10.1016/j.edurev.2017.12.001

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. doi:10.1037/h0047358

DePaoli, J. L., Atwell, M. N., & Bridgeland, J. (2017). *Ready to lead: A national principal survey on how social and emotional learning can prepare children and transform schools*. Washington, DC: Civic Enterprises. Retrieved from http://www.casel.org/wp-content/uploads/2017/11/ReadyToLead_FINAL.pdf

DeRosier, M. E., & Thomas, J. T. (2019). Hall of heroes: A digital game for social skills training with young adolescents. *International Journal of Computer Games Technology, 2019*, Article ID 6981698. doi:10.1155/2019/6981698

Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gullotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. New York, NY: Guilford Press.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. [doi] doi:10.1111/j.1467-8624.2010.01564.x

Dusenbury, L., Dermody, C., & Weissberg, R. P. (2018). *State scorecard scan*. Chicago, IL: Collaborative for Academic Social and Emotional Learning. Retrieved from https://casel.org/wp-content/uploads/2018/09/csi-scorecard-sept2018.pdf

Edge Research. (2018). *Developing life skills in children: A road map for communicating with parents*. Arlington, VA: Author. Retrieved from https://bealearninghero.org/parent-mindsets/

Gallup. (2018). *Assessing soft skills: Are we preparing students for successful futures?* Washington, DC: Author. Retrieved from https://www.nwea.org/content/uploads/2018/08/NWEA_Gallup-Report_August-2018.pdf

Gresham, F. M., & Elliott, S. N. (2017). *Social skills improvement system social-emotional learning edition*. Bloomington, MN: NCS Pearson.

Jagers, R. J., Rivas-Drake, D., & Borowski, T. (2018). *Equity and social and emotional learning: A cultural analysis*. Chicago, IL: Assessments Work Group. Retrieved from https://measuringsel.casel.org/wp-content/uploads/2018/11/Frameworks-Equity.pdf

Jones, S. M., & Doolittle, E. J. (Eds.). (2017). Special issue on social and emotional learning. *The Future of Children*, *27*, 3–178.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-second edition (NEPSY-II)*. San Antonio, TX: Pearson Assessment.

Krachman, S. B., & LaRocca, B. (2017). *The scale of our investment in social-emotional learning*. Boston, MA: Transforming Education. Retrieved from https://www.transformingeducation.org/wp-content/uploads/2017/09/Inspire-Paper-TransformingEd.pdf

LeBuffe, P. A., Shapiro, V. B., & Robitaille, J. L. (2018). The Devereux Student Strengths Assessment (DESSA) comprehensive system: Screening, assessment, planning, and monitoring. *Journal of Applied Developmental Psychology*, *55*, 62–70.

McKown, C., & Taylor, J. (Eds.). (2018). Social-emotional assessment to guide educational practice [Special issue]. *Journal of Applied Developmental Psychology*, *55*(March–April). doi:10.1016/j.appdev.2017.12.002

McKown, C. (2017a). Promises and perils of assessing character and social emotional learning. *Journal of Character Education*, *13*, 47–52.

McKown, C. (2017b). Social-emotional assessment, performance, and standards. *The Future of Children*, *27*(1), 157–178. doi:10.1353/foc.2017.0008

McKown, C. (2019). *Assessing students' social and emotional learning: A guide to meaningful measurement (SEL solutions series)*. New York, NY: W.W. Norton.

McKown, C., Gumbiner, L. M., & Johnson, J. K. (2011). Diagnostic efficiency of several methods of identifying socially rejected children and effect of participation rate on classification accuracy. *Journal of School Psychology*, *49*(5), 573–595. doi:10.1016/j.jsp.2011.06.002

McKown, C., Russo-Ponsaran, N., Johnson, J. K., Russo, J., & Allen, A. (2016). Web-based assessment of Children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, *34*(4), 322–338. doi:10.1177/0734282915604564

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*(2), 498–515. doi:10.1111/1467-8624.7402012

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, *46*(3), 235–261. doi:10.1016/j.jsp.2007.05.001

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing you children's views of their academic, social, and emotional lives: An evaluation of the Berkeley Puppet Interview. *Child Development*, *69*, 1556–1576.

Merrell, K. W. (2009). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York, NY: Routledge.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. doi:10.1037/0003-066X.50.9.741

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge/Taylor & Francis Group.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, *17*(4), 297–334. doi:10.1177/014662169301700401

Molnar, M. (2018). *Money is flowing to social-emotional learning: Allstate foundation dedicates $45 million*. Bethesda, MD: Education

Week Market Report. Retrieved from https://marketbrief.edweek.org/marketplace-k-12/money-flowing-social-emotional-learning-allstate-foundation-dedicates-45-million/

RAND. (n.d.). RAND Education Assessment Finder. Retrieved from https://www.rand.org/education-and-labor/projects/assessments.html

Schonert-Reichl, K. A., Hanson-Peterson, J. L., & Hymel, S. (2015). SEL and preservice teacher education. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning; research and practice* (pp. 406–421). New York, NY: Guilford Press.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the preschool self-regulation assessment (PRSA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173–187. doi:10.1016/j.ecresq.2007.01.002

Spivack, G., & Levine, M. (1964). Consistency of individual differences in time judgments. *Perceptual and Motor Skills*, *19*(1), 83–92. doi:10.2466/pms.1964.19.1.83

Steinberg, S. (1976). *View of the world from 9th Avenue*. New York, NY: The New Yorker.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. doi:10.1111/cdev.12864

Wan, T. (2017). *Panorama education raises $16M to connect emotional, academic wellbeing with data*. Burlingame, CA: EdSurge.

Weissberg, R. P., Goren, P., Domitrovich, C., & Dusenbury, L. (2012). *Effective social and emotional learning programs: Preschool and elementary school edition.* Chicago, IL: Collaborative for Academic Social and Emotional Learning. Retrieved from https://casel.org/preschool-and-elementary-edition-casel-guide/

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2018). Development and implementation of student social-emotional surveys in the CORE districts. *Journal of Applied Developmental Psychology*, *55*, 119–129.