

Are You READY TO ASSESS Social and Emotional Development?

Brief



STOP

Think

Act

Why Now?

Education in the 21st century is awakening to a call for students who are not only proficient with academic content but who also have developed the social and emotional (SE) knowledge, attitudes, and skills that are necessary for success in college and careers. Throughout this brief, we refer to that combination of knowledge, attitudes, and skills as SE competencies. SE competencies enable students to harness their academic knowledge and turn that knowledge into action. Research suggests that the development of academic knowledge and SE competencies is inextricably linked and the two elements mutually reinforce one another (Farrington et al., 2012). Leaders in education have recognized the critical need to support students developing SE competencies and, increasingly, have turned their attention to assessments to ensure accountability to standards that emphasize them (Dusenbury, Zadrazil, Mart, & Weissberg, 2014). Although this shift in focus toward a more holistic approach to student success may be a significant and positive advance for education, it begs certain questions about policy and practice. This brief and the accompanying *Ready to Assess Decision Tree* and *Tools Index* aim to provide policymakers and leaders in education with (1) an overview of the developing assessment landscape and (2) a framework and guidance for deciding if and when you are ready to assess SE competencies.

Assessment Landscape

Educational assessment has grown at a rapid pace during the past 20 years in its scope, sophistication, intensity, and most notably, its frequency within the classroom. Beginning with the increased focus on academic content standards brought about by the 1994 reauthorization of the Elementary and Secondary Education Act and culminating with the demands for standardized testing for accountability purposes emphasized by No Child Left Behind, high-stakes academic assessments have become a fixture of the K–12 classroom. Between state- and district-mandated assessments, students participate in as many as 20 tests per year (Lazarin, 2014). Although these assessments have brought accountability to content standards and equity in education to the forefront, the varying quality of the many standardized tests students are taking, and even states' content standards themselves, have been called into question (Yuan & Le, 2012).

To ensure a high level of, and consistent quality for, state academic content standards—and the standardized tests associated with them—most states have adopted the Common Core State Standards, which provide a national benchmark for all students. At the same time, test developers have expanded the scope of what can be assessed in response to calls for a more nuanced approach to promoting and evaluating student's SE competencies. This shift has resulted in a new wave of educational assessments that aim to uphold accountability to the more rigorous academic content standards (i.e., Partnership for Assessment of Readiness for College and Careers and Smarter Balanced), as well as a push to use additional, alternative assessments to measure SE competencies (Soland, Hamilton, & Stecher, 2013).

Assessments that measure SE competencies range from comprehensive to specific. For example, the Academic Competence Evaluation Scales measure students' interpersonal skills, motivation, engagement, study skills, and academic skills (DiPerna & Elliott, 1999). Other assessments, such as the Behavior Intervention Monitoring Assessment System, focus more narrowly on SE competencies reflected by student social adjustment and behavior (McDougal, Bardos, & Meier, 2011). We present a sample of these assessments in our *Ready to Assess Tools Index*.

Assessing students' SE competencies could yield many benefits for students, educators, and policymakers if assessments are implemented efficiently and effectively. The integration of assessments for SE competencies with those for academic content standards could promote a greater awareness about the critical role that SE competencies play in the classroom, as well as in postsecondary education and careers. This heightened awareness might lead to improvements in policy and practices that promote the development of the SE competencies that well-rounded students need to succeed. However—especially considering the existing load of assessments for academic content standards students already face—certain precautions and considerations must be taken into account before adding another set of assessments to the mix. We review those precautions and considerations in the sections that follow, which serve as the touchstones of our *Ready to Assess Decision Tree* and *Tools Index*.

Ready to Assess?

Existing state and district assessment requirements, as well as the sheer novelty of assessments for SE competencies, raise basic concerns about schools' capacity to implement SE assessments and properly interpret and act on their results. Rolling out these cutting-edge assessments without complementary teacher and staff professional development, as well as systemic capacity building, and processes to produce rigorous, valid and reliable test results, may lead to incomplete or even troublesome results for educators and students. Policymakers and education leaders can avoid these pitfalls by carefully considering the four key Ready to Assess elements: purpose, rigor, practicality and burden, and ethics.

Purpose

The decision to use any assessment must be grounded by a clear and well-founded purpose. Although it might seem obvious at first, the purpose(s) of an assessment may be highly nuanced and can easily become confounded by competing interests, mixed messaging, or mismatched expectations. Three common assessment purposes are (1) accountability, (2) communication, and (3) information.

Accountability

Assessments chosen for accountability purposes may vary depending on the audience to whom accountability is due and the implications or consequences for meeting established requirements. Policymakers concerned with accountability to state standards (including social and emotional learning standards) and establishing funding for various programs face a high-stakes effort that may demand the most rigorous evaluations, requiring large-scale, standardized assessment. Educators and practitioners concerned with demonstrating more local impact or achieving school mission objectives may face lower stakes and be better served by smaller-scale, customizable assessments.

Communication

Educators and policymakers communicate regularly with diverse audiences and for many different purposes. At the state, district, and school levels, educators and policymakers may need supporting evidence to make their case for a need, to satisfy a request for information, or to communicate with external stakeholders

(e.g., parents, industry, and the community at large). In that case, the type of assessment will vary with the audience and intended outcome. Demonstrating compelling evidence to advocate for or against a given policy may be relatively high-stakes, and require large-scale, highly rigorous assessments. Smaller-scale, tailored assessments can be used to meet lower-stakes, isolated requests for information or explanation, or they can be used for “storytelling” purposes.

Information

One of the great promises of using assessment data is that the information will contribute to well-informed decision making at the local, state, and federal levels. Gathering information using assessments can serve many purposes: It may be broadly exploratory, or it can inform a need for changes in policies or practices; it can help improve performance or practice by providing formative feedback to students and practitioners, build capacity, or determine further PD efforts; and it can provide proof of the effectiveness of policy or practice. Finally, on a more basic level, assessments can be used to identify student needs for intensive support (e.g., individualized education programs), to provide intentional instruction, or to guide intervention.

STOP

After considering the many purposes that assessments for SE competencies might serve and then choosing one or more as the basis for using an assessment, stop and consider whether the purpose necessitates assessment. Although using an assessment may be highly appealing because of its potential power in validating results, often student, school, district, or state data already exist to meet these and many other needs. It is worthwhile to ask whether more data are needed or whether the need can be addressed in alternative ways, for instance by using existing data and assessments or by using a proxy for youth outcomes such as assessing elements of school climate including the quality of teacher practice and other environmental factors supporting SE development. Finally, at each stage of the decision process, consider the risks and benefits of pursuing these goals with assessment methods. If assessment is the best option, then the next consideration in the Ready to Assess framework is rigor.

Rigor

After firmly establishing the rationale for using an assessment and the stakes involved, determining the rigor of the prospective assessment is of the utmost importance. We encourage practitioners to rigorously implement any assessment under consideration; however, when considering the term, “rigor,” in the *Ready to Assess Tools*, we are referring to how comprehensive the assessment is and the degree to which the assessment is a well-established, valid, and reliable measure of SE competencies. The first dimension of rigor to consider is the assessment type, which can vary depending on whether the purpose chosen is relatively high stakes or low stakes.

Lower-rigor assessments may be appropriate for lower-stakes purposes, such as information gathering and communication—especially at the local level, for a single school or program. Higher-rigor assessments may be required for higher-stakes accountability purposes, especially at districtwide and statewide levels of reporting. However, in some cases, the stakes involved may not match with the desired level of rigor. It may be appropriate, or even highly

Lower Rigor

- Authentic assessment (portfolios, play based, and journals)
- Observation
- Interview
- Self-report or survey (homegrown or miniature version)

Higher Rigor

- Performance
- Self-report or survey (valid, reliable, normed, and widely available)
- Observation
- Interview

desirable, to include some lower-rigor assessments for high-stakes purposes, especially if those assessments are implemented rigorously, and if they provide depth to the assessment results. At the same time, in low-stakes situations, higher-rigor assessments can be used effectively if certain practical implementation considerations—such as capacity and program maturity—can be ensured.

The rigor of the assessment depends largely on its levels of validity and reliability. Ideally, the assessment under consideration has a record of successful use by other state and local actors, for similar purposes. This will help stakeholders determine the significance of the assessment results within the larger picture of all other education data. An assessment should have high levels of reliability and validity to be considered for use on a large scale or for decision making. In some cases, multiple assessments or assessment types may be necessary to achieve the established purposes.

After all the elements of rigor are determined to be sufficient for the purposes identified, it is critical to develop a theory of action and a concrete plan for how to use the assessment results. Ultimately, this plan will determine whether even highly significant and meaningful results can be used to serve the intended purpose of the assessment. For instance, if the assessment does not have strong evidence of validity and reliability, then this should be clearly communicated and the results should be considered exploratory rather than final or definite. Conversely, the decision to assess at all may be postponed until higher levels of validity and reliability can be ensured.

Validity: The degree to which an assessment can be said to measure what it is supposed to be measuring (e.g., communication skills), instead of something else (e.g., mathematics ability).

Reliability: The degree to which an assessment can be expected to produce the same results after being administered multiple times to the same population.

A **theory of action** is an “if, then” statement that articulates the mechanisms by which the desired outcomes will be achieved via the selected means. In the case of using assessments of skills and social and emotional competencies, a theory of action should clearly state how using the assessment—including its implementation and the analysis of results—will lead to achieving your identified purposes.

THINK

After reviewing the key elements of rigor, think critically about how the rigor of your chosen assessment maps onto your purposes. It may become clear that there is not a good match between the level of rigor of the available assessments and your intended purpose and outcomes. In some cases, you might be able to take less formal measures to achieve the same outcomes, such as teacher reports, parent-teacher conferences, out-of-school time supports, or routine counseling or other services. If these options are not adequate and the rigor of your chosen assessment matches the purpose you have identified, then the next step is to consider practicality and burden.

Practicality and Burden

Having developed a clear sense of purpose and the rigor of the assessment, outlining the relevant practical considerations and estimating burden—or implementation costs—is paramount. Even with the most clearly articulated purpose and correspondingly rigorous assessments, a disconnect between those factors and real-world practicalities and burdens could derail your process.

Practicality

Consider two key practical considerations before estimating burden: (1) the age of the program or initiative in question and (2) the number of youth served. Each of these factors may vary independently from one another, so it is important to consider the possible permutations between them. On the one hand, in the case of new programs with a small number of youth, it may not be feasible to implement high-rigor assessments. Beyond feasibility, high-rigor assessment results may not be maximally useful when applied to small groups or with

little opportunity to reveal change across time. In such cases, it may be more appropriate to use a lower-rigor assessment or to hold off on assessing altogether until the program or initiative is more mature and a greater number of youth can be evaluated. On the other hand, an assessment of a mature program with a large number of youth may yield a much greater return on investment when using a high-rigor assessment instead of implementing a relatively lower-rigor evaluation only to save on costs. These general guidelines follow from the fact that high-rigor assessments usually have higher levels of validity and reliability but are more complex and costly to implement, whereas lower-rigor assessments are less likely to have well-established and high levels of validity and reliability, but are much less complex and costly to implement (Soland et al., 2013). Ensuring a proper articulation of purpose, rigor, and practicality will allow for the most accurate and productive comparison of potential benefits with implementation costs (or burden).

Burden

Elements of burden include staff capacity, infrastructure requirements, data use, budget, and risks to teachers, staff, students, and families. These barriers to implementation can limit the maximum return on investment already identified via the establishment and articulation of purpose, rigor, and assessment practicality. It is important to consider each element of burden carefully before continuing with assessment plans.

Ensuring staff capacity requires training staff on assessment administration and data collection, data analysis, and reporting—or contracting out those services to another agency. When evaluating infrastructure, it is important to consider mechanisms for data collection (e.g., computerized assessments), data storage, and data analysis tools. The use of assessment data requires a plan for data analysis, as well as the continued use of that data and data from future assessments. All of these involve unique costs, both monetary and in terms of labor and time. Finally, consider the burden on teachers, students, and families, given the implementation costs and identified capacity as potential barriers to achieving maximum return on investment. After weighing the burdens you have identified against the potential benefits of assessing, the final checkpoint before acting is a consideration of ethics.

Ethics

Once the purpose, rigor, practicalities, and burden of using an assessment have been established and a plan for its rollout has been developed, it is time to stop and think before acting—to do a final check for ethics and consider the big picture. At this time, convene your team members and evaluate how the use of the assessment and the associated potential risks and benefits fit with the larger mission of the institution and community. Ask yourselves: Does your purpose require a high-rigor assessment? Does the benefit of getting the data outweigh the risks to participants, and are you administering assessments to a group already burdened with surveys and tests? Consider any other options to reduce the risk and maximize the benefit of the outlined plan. Use the *Ready to Assess Decision Tree*—which follows this narrative outline—to help ensure that after deeply considering these four components, the decision to use (or not use) an assessment, is of greatest potential value.

ACT

After fully mapping out the connections between your assessment purpose, rigor, and burden, and determining that the potential value of the assessment in question matches up with your ability to implement it, ethically, and at a relatively low cost, it is time to take action. The *Ready to Assess Decision Tree* and *Tools Index* can be a handy reference, even as you begin to implement or refine your plan. Keeping all of the above considerations in mind throughout the process of developing and implementing your assessment plan can help to ensure that assessment goals truly meet education needs.

References

- DiPerna, J. C., & Elliott, S. N. (1999). The development and validation of the Academic Competence Evaluation Scales. *Journal of Psychoeducational Assessment, 17*, 207–225.
- Dusenbury, L., Zadzil, J., Mart, A., & Weissberg, R. (2011). *State learning standards to advance social and emotional learning: The state scan of social and emotional learning standards, preschool through high school*. Washington, DC: Collaborative for Academic, Social, and Emotional Learning.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.
- Lazarin, M. (2014). *Testing overload in America's schools*. Washington, DC: Center for American Progress.
- McDougal, J. L., Bardos, A. N., & Meier, S. T. (2011). *Behavior Intervention Monitoring Assessment System Technical Manual*. Toronto, Canada: Multi-Health Systems.
- Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st century competencies: Guidance for educators*. Washington, DC: RAND Corporation.
- Yuan, K., & Le, V. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests*. Washington, DC: RAND Corporation.