

Also Available

The Early Education Leader's Guide:
Program Leadership and Professional Learning
for the 21st Century

*Nonie K. Lesaux, Stephanie M. Jones,
Annie Connors, and Robin Kane*

Making Assessment Matter:
Using Test Results to Differentiate Reading Instruction

Nonie K. Lesaux and Sky H. Marietta

Teaching Advanced Literacy Skills:
A Guide for Leaders in Linguistically Diverse Schools

*Nonie K. Lesaux, Emily Phillips Galloway,
and Sky H. Marietta*



Measuring Noncognitive Skills in School Settings

Assessments of Executive Function
and Social–Emotional Competencies

edited by

Stephanie M. Jones
Nonie K. Lesaux
Sophie P. Barnes

Foreword by Timothy P. Shriver



THE GUILFORD PRESS
New York London

● CHAPTER 2

The Fidget Spinner Effect

Social and Emotional Assessment and the Healthy Evolution of the Social and Emotional Learning Field

Clark McKown

Social and emotional learning (SEL) refers to the competencies children use to interact successfully with each other, to participate constructively in group activities, and to form and deepen relationships. SEL includes competencies like self-awareness, self-management, social awareness, relationship skills, and responsible decision making (www.caseli.org). We know from extensive research that the better students' social and emotional competencies are developed, the better they do in relationships, in school, and in life (McKown, 2017).

Social and emotional competencies are also teachable: Done well, structured curricula that focus on explicit instruction of student social and emotional competencies (referred to in this chapter as "SEL programs") can produce desirable outcomes as diverse as improved behavior, mental health, and academic outcomes (Mahoney, Durlak, & Weissberg, 2019). The best evidence of SEL's promise comes from well-constructed experimental or quasi-experimental field trials (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Outside of explicit instructional programs, we know that the quality of teacher-student interactions and children's relationships with their teachers (referred to here as "SEL practices") have long-term academic, social, and emotional consequences (Hamre & Pianta, 2001). It is this foundation in rigorous research that places SEL in the realm of the "evidence-based." Although SEL programs and SEL practices are both key components of school-based SEL initiatives, the focus of this chapter is on SEL programs, and the role of assessment in supporting their effective use.

Evidence that SEL programs are beneficial has ushered in a new era in which SEL programs are being adopted at a rapid clip. In two national

surveys, 70% or more of elementary principals indicated that they had developed or implemented a plan to address student social and emotional competence. More than half reported that they had adopted an SEL program or curriculum (Atwell & Bridgeland, 2019; Hamilton, Doss, & Steiner, 2019).

End of story? Not by a long shot.

One of the key challenges that will soon face the field is the same challenge facing any field in which an evidence-based practice (EBP) is adopted on a massive scale. Specifically, how do we ensure that the SEL programs used in a large and growing number of schools in the United States contain the elements that made them effective in carefully executed studies such as those summarized in influential meta-analyses? On the one hand, we have learned that evidence-based SEL programs, implemented with sufficient intensity and quality, support positive student outcomes. On the other hand, if we assume that with the meta-analyses, we have all the evidence we need, and no more data are needed to scale up programs with fidelity and effectiveness, it is highly unlikely that SEL will maintain quality and impact at scale.

Without some way of measuring what is happening and whether it is making a difference, the field may find itself in trouble. What might happen, you ask? SEL programs that are effective in the context of well-resourced studies could be implemented too inconsistently or with insufficient fidelity to produce a benefit. I have spoken with many school administrators who reported that their district adopted an SEL program, and whether, how much, and how well that curriculum is being used varies from one classroom to the next. Many of those administrators do not know precisely who is using the curriculum or how well, so they do not know where to direct instructional coaching resources. Without assessment—of implementation, for example—it is impossible to know.

There is another problem that can arise from going to scale without assessment. Practices that are not known to benefit teaching, learning, or student outcomes may be marketed as SEL and adopted without evidence that they are beneficial. This is not always a problem because some well-designed programs may in fact benefit students even though a rigorous study has not yet demonstrated their efficacy. If educators use assessment—of student competencies, for example—they can monitor whether and how much students are acquiring the skills an "unproven" program is designed to nurture. However, unless educators assess student social and emotional growth, it is impossible to distinguish beneficial programs from ineffective programs.

When effective programs are not implemented well, and ineffective programs are marketed as evidence-based SEL, as the field grows, and student outcomes do not improve, this could lead to the erroneous conclusion

that SEL is a waste of time, money, and energy. Critics will say SEL does not work when in fact, it is *ineffective* SEL that does not work. The risk of either or both of these problems increases with scale. Assessment can help educators use data to guide practice, measure progress, and ultimately, maintain the integrity and impact of SEL programs at scale. With assessment, educators can know how much and how well they are using an evidence-based program as intended, and how much those actions are contributing to student competence and other important outcomes. With assessment, educators can also know how much promising but unproven programs are helping.

• A Study in Contrasts

What does this all look like on the ground? Let's consider three teachers who heartily agree that SEL matters and who report they are engaged in practices that support student social and emotional development. One teacher's district adopted an evidence-based SEL program after reviewing a number of options, including reviewing evidence of their efficacy. District and school leadership strongly support SEL and expect all educators to use the program, and use it well, and they provide resources to support effective program use. The teacher has participated in professional development workshops with the program provider and ongoing structured coaching from the district to learn to use the program resources proficiently. Her goal is to teach students the social and emotional competencies they will need to succeed in school and life. At the beginning of the year, she administers a benchmark social and emotional competence assessment to all students. Each day, she teaches SEL lessons, following the scope and sequence, but adapting the lessons expertly as opportunities to connect to student experiences permit. During literacy, she reinforces social and emotional competencies by integrating intentional probes and questions. The benchmark assessment reveals that many of her students struggle with social problem solving, a key skill in conflict resolution. To address this need, she provides students opportunities to problem-solve hypothetical interpersonal conflicts. During literacy, when there are conflicts between characters, she encourages students to think through the nature and possible resolutions of the problem. Through her questioning during conflict resolution exercises and literacy, she formatively assesses student progress. Periodically, she administers a more formal assessment to measure student progress in response to her efforts.

A second teacher works in a district that has purchased an SEL program that was identified by a cabinet-level administrator at a conference. After a surprise announcement, the teacher participated in a 1-day workshop on using the program. No one has communicated expectations for

program use. After the workshop, district and school leaders rarely mention the program. The teacher sees the SEL program as "one more thing" she has to do. After 3 months, in response to some social conflict between students, she reviews the first lesson plan, having forgotten most of what was covered in the workshop. She teaches the lesson but it does not go well and she decides that the program does not work and does not use it again.

A third teacher is concerned about student misbehavior. Several of his students are frequently off-task, out of their seats, and impulsive. His goal is to reduce the frequency of problem behaviors in these students. He heard from a colleague that fidget spinners support self-management, and this is his primary SEL "intervention."¹ Fidget spinners are a handheld manipulative widely marketed as a way to support greater focus. To support student self-management, he provides fidget spinners to his students who he has observed most frequently interrupting or engaging in off-task behavior. The students spin their fidget toys frequently during class, and other students tell the teacher they find them distracting. The teacher informally observes all students' behavior and develops an overall impression of how things are going. When asked by a colleague whether the fidget spinners work, he indicates his impression is that he is not sure, but he thinks so. He indicates that grades and test performance among the students who use the fidget spinner have not changed, but his overall impression is that they are less disruptive.

I would wager students in the first teacher's class will benefit from her efforts. The first teacher uses a program for which there is evidence of efficacy. She has engaged in systematic training on how to use the program skillfully. She uses assessment to guide SEL instruction for all students. Throughout the day, she creatively integrates multiple SEL touchpoints so the whole class can practice social and emotional competencies, particularly in areas of need identified in a whole-class assessment. In that way, SEL becomes part of the fabric of instruction. Will these efforts result in student academic, social, and emotional development? There are of course no guarantees. However, she is using approaches that have been shown to work and she is using them skillfully, with data and professional learning supports to guide her. In addition, she is measuring student growth, so she will be able to use data to see how much her students are acquiring social and emotional competencies.

I would also wager that students in the second teacher's class will not benefit from the SEL program, because it barely made it out of the proverbial shrink wrap.

My final bet: Students in the third teacher's class will benefit minimally, if at all, from his "SEL" intervention. The third teacher is using a very specific tool to reduce problem behaviors in a small number of students. His social and emotional goal is not to build strengths, but to

¹Fidget spinners are not evidence-based.

extinguish problems. There is no evidence that fidget spinners can accomplish this goal beyond word of mouth and anecdotes. Outside of the fidget spinner, he does not change instruction to support the self-management skills of students, and he receives no professional-learning support to guide his practice. Furthermore, he has little evidence to know whether students' use of the fidget spinner changed behavior. He does, however, have evidence that other students find them distracting.

• **DWI: Doing (SEL) without (Good) Information**

The problem here is that all three teachers call what they are doing "SEL," but their approaches are very different in their likelihood of benefiting students. As the field grows, it seems likely that many activities will be described as "SEL," a phenomenon I refer to as "the fidget spinner effect." Although the fidget spinner example is rather extreme, it is highly likely that unproven but seemingly credible (often more credible than fidget spinners) SEL initiatives and effective but poorly implemented SEL approaches could also all be called "SEL." As SEL begins to unwittingly assimilate ineffective practices, and those practices predictably produce no measurable benefit, educators and the public could easily dismiss SEL as another fad. The proverbial baby could easily go the way of the bathwater.

To be clear, educators need a great range of tools to address their students' needs and they will need to adapt "out-of-the-box" programs to the local community—programs reflect only one such tool. They may, for example, adopt practices such as restorative justice or circle time or other kernels of social and emotional practice (Jones, Bailey, Brush, & Kahn, 2017) that reflect adult behaviors and processes rather than structured programs, and that is an important part of what educators should do. The key point is that when an educator implements a program or a practice with the aim of producing a social and emotional benefit, it is preferable that the program or practice has been demonstrated in the context of rigorous research to achieve the intended social and emotional benefit. In addition, whether or not such evidence exists, it is critical that educators use assessment to guide programs and practices, and measure the outcomes those programs and practices produce. In doing so, they can determine what is working, and let go of what is not.

• **Bringing Evidence-Based Programs to Scale (and Maintaining Impact)**

The field of education, and SEL within it, is not alone in struggling with the challenge of bringing evidence-based practices (EBPs) to scale, by which

I mean taking practices that were originally used in a limited number of locations and implementing them in a massive number of settings. This is happening with SEL—as mentioned previously, more than half of U.S. school districts are implementing some form of SEL program. Right now, SEL programs are going to scale, often without a coordinated effort to assess social and emotional learning.

We can learn a great deal from other fields that have worked to bring evidence-based programs to scale. Alongside education, fields such as behavioral health, child welfare, juvenile justice, and public health all face the challenge of bringing evidence-based programs to scale (Fagan et al., 2019). Recognizing this challenge, the Society for Prevention Research (SPR) convened a task force to consider ways to bring evidence-based programs to scale in each of the five public systems mentioned. The rationale for doing so is perhaps obvious, but so it does not go unstated, bringing effective programs to scale in public systems has great potential to support population health.

The SPR task force identified specific factors that affect the scale-up of EBPs in public systems. One key factor is "statutory endorsement," which includes laws and regulations supporting the use of EBPs, and the flexibility to use funds to support their adoption. In that, the field of SEL has made strides, with 18 states (and counting) having adopted policies that indicate what social and emotional competencies students should know and be able to demonstrate (Dusenbury, Yoder, Dermody, & Weissberg, 2020). Another critical component to scaling effectively includes what the task force described as "data monitoring and evaluation capacity." That includes the routine collection of data on implementation and the outcomes EBPs are expected to produce and the use of those data to engage in practice improvements.

What should be assessed? In my view, educators will ideally assess SEL program implementation, student social and emotional competence, and climate. *Implementation* refers to the extent to which educators are using an SEL program as intended, and with sufficient intensity and quality. *Student competence* refers to the social and emotional knowledge, skills, and dispositions that are the targets of instruction in SEL programs, and that are described by the Collaborative for Academic Social and Emotional Learning (www.casel.org). *Climate* refers to the conditions of learning as students perceive it, and includes things like how safe students feel, how connected they feel to adults and peers, and how supported they feel as learners (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). By assessing these three elements, educators can know what adults are doing, and how those actions are affecting student competencies and the conditions of learning. Assessing program implementation helps educators direct support for high-quality program use where it is needed most. Assessing student social and emotional competence early in the year can help guide

what to teach to whom; assessing competence after a period of instruction can quantify growth. Similarly, assessing climate can help adults take actions to create a warm and inclusive learning environment, and monitor whether those actions are improving the climate.

In practice, what might an assessment system look like that reflects the kind of data monitoring and evaluation needed to support SEL at scale? My colleagues and I have developed one such system. We started by developing SELweb, a performance-based direct assessment of student social and emotional competencies for the elementary grades (McKown, Allen, Russo-Ponsaran, & Johnson, 2013). SELweb is a nationally normed web-based application that includes illustrated and narrated media that takes about 30 minutes for children to complete. Children interact with a series of tasks in which they demonstrate their ability to read others' feelings and perspectives, to solve social problems, and to manage their emotions. It is not a traditional self-report survey in which children rate their own competencies—rather, by using performance tasks, children have to show what they know and understand. Educators use SELweb to benchmark skills early in the year, make decisions about SEL instruction, and measure progress after a period of instruction.

Recognizing that a complete data-monitoring and evaluation system requires more than student competence assessment, we recently added a brief student climate survey measuring students' perceptions of safety, belonging, and instructional support. In addition, we added a brief teacher SEL program adherence survey to measure program implementation. SELweb is designed to measure dimensions of competence that are commonly the targets of instruction in SEL programs, and so it reflects a program-agnostic approach to measurement. Districts can flexibly adopt the assessment components that meet their needs, and can measure implementation, competence, and climate within a simple and unified assessment and reporting system.

SELweb, and other emerging SEL assessment systems (Assessment Work Group, 2019; Thapa et al., 2013), give hope that if the field moves fast, it can maintain the kind of commitment to data that elevated it to prominence in the first place, as reflected in meta-analyses and the studies they summarized (Durlak et al., 2011; Mahoney et al., 2019). This time, however, assessment data are not gathered and presented to establish initial evidence of efficacy (see meta-analyses). Instead, assessment data collection and use becomes an integral part of SEL program use—educators use assessment data to benchmark student competence and climate, to decide what to do to foster student competencies and create a positive climate, to monitor what teachers are doing so that coaching resources can be deployed based on need, and to measure change in student competence and climate. By intentionally collecting and reviewing strategically selected

assessment data, educational decision makers—from superintendents to classroom teachers—have information about what is happening, how well it is happening, and what outcomes are associated with these activities. The regular availability of assessment data supports cycles of reflection and practice improvement.

The field of SEL will benefit greatly from adopting usable and feasible methods to measure SEL program implementation and the social-emotional competencies and dimensions and climate factors those programs are intended to influence. Building on educational traditions of data use, what if the field of SEL could measure SEL program implementation and outcomes? What would happen if policymakers, educators, and program developers committed to implementing assessment systems to support high-quality SEL program implementation? What might that look like? What should be assessed? How should the data be used? Next, I consider two models for the SEL field to consider. The first model, improvement science, is a method of conducting cycles of assessment, reflection, and action to address locally defined programs of practice. The second, positive behavioral interventions and supports, or PBIS, is a school-based systems intervention designed to support reductions in problem behaviors and increases in positive behaviors. Notably, PBIS integrates assessment and reflection on data to continuously improve practice.

Model 1: Improvement Science

What It Is

Improvement science (Bryk, Gomez, Grunow, & LeMahieu, 2017) offers a helpful framework for considering how assessment data might be used to support the healthy evolution of the field. This framework makes several important assumptions about systems change that can be instructive to efforts to scale up SEL. First, it assumes that systems change requires a strong understanding of the system, its routines, work processes, resources, and demands. Second, it assumes that to succeed in efforts to change systems requires input from the participants in the system. Third, it implies that systems change is essentially a locally managed phenomenon—in the case of scaling SEL, for example, the needs and pressures in which the SEL initiative is to be executed will vary from district to district and school to school. From an implementation science perspective, to succeed, local needs must be understood, accounted for, and addressed. Fourth, it assumes that successful change efforts will create and use a model or theory of the factors that need to be addressed to produce change in work processes (such as SEL implementation) and the outcomes they produce (such as student social and emotional competence). Finally, improvement science assumes

that those in the system, in collaboration with experts in the field and others, should use data to monitor intended (and unintended) changes in work processes and the outcomes they are designed to produce. Data should be regularly collected to show successes and failures. Failures are acceptable as long as they guide modifications to the systems change effort and continued testing of those efforts. Improvement science is a way to clearly define local problems of practice and develop a rigorous plan for iteratively testing and evaluating approaches to solving the problem. A key premise of implementation science is that by developing work processes focused on data-informed cycles of practice improvement, positive systems change is more likely to take root.

What the Field of SEL Can Learn

What can those of us who want to see SEL scale with quality learn from improvement science? First, understanding what problem SEL programs solve for educators—from the point of view of educators—is key. When, for example, educators see an SEL program as a compliance requirement handed down from district leadership or beyond, the level of motivation and interest in implementation may be different than when they see an SEL program as a core element of their professional identities and a central driver of student success. An improvement science perspective might therefore suggest that success at scale requires us to assess and address educators' level and kind of motivation to adopt SEL programs.

Second, improvement science suggests that it is important to understand and account for local contexts, particularly the resources, work processes, and pressures endemic to a particular setting. For example, a district that has launched several new curricular initiatives may be at a different level of readiness to implement an SEL program than one that has been studying SEL, socializing its work force, and otherwise preparing to launch for 2 years. A district that has well-established data systems may be in a better position to use SEL assessment to support implementation than one that is grappling with data use. Broadly, therefore, an improvement science perspective might lead us to systematically assess readiness, capacity, and competing pressures within a district and plan implementation in a way that accommodates these factors. Such an assessment would ideally lead to a theory of the key drivers that influence the degree to which SEL programs can be successfully and deeply integrated into practice.

Third, an improvement science perspective suggests that the formulation of key drivers should guide targeted efforts to change the system. For example, a district might identify the following drivers of consistent high-quality SEL program implementation: (1) principal leadership and advocacy for SEL programs, (2) teacher skill in teaching the curriculum,

(3) teacher belief that SEL is important, (4) time for planning and instruction, and (5) teachers' own social and emotional wellness. Each of these hypothesized drivers has constituent elements that can be easily operationalized and measured. For example, teacher social and emotional wellness might include: (1) teachers' level of job-related burnout, (2) teachers' stress-management knowledge and skill, and (3) perceived support from principals. District leadership might decide that this is the most important driver to focus on and that they will invest in teacher wellness workshops and supporting positive teacher-principal relationships. The broad hypothesis might be that improvements in teacher wellness will lead to increased belief that SEL is important, paving the way for implementation.

Fourth, improvement science suggests that a district's hypotheses about the drivers are testable and therefore possibly wrong, and the district's decision to focus on a particular driver is an experiment. To test and refine the district's strategy requires assessment related to the strategy in question. In this case, that might involve assessing: (1) attendance at wellness workshops, (2) teacher ratings of workshop quality, (3) teacher and principal ratings of whatever intervention is designed to improve teacher-principal relationships, (4) teacher wellness as defined above, and (5) teacher beliefs about the importance of SEL. These data would be gathered at the onset of the initiative and frequently enough throughout to monitor what is happening, how well it is going, and whether the key outcome of interest (teacher attitudes about SEL) are moving in response. Where the data suggest that something is not working, this might precipitate a change in strategy.

Note that the above example did not include any assessment of student social and emotional competence. This highlights the importance of considering multiple forms of SEL assessment that can accomplish a variety of goals, all in service to consistent and high-quality programs. Student competence assessment is a key tool in that portfolio, along with measures of classroom practices. From an improvement science perspective, classroom practice is a key hypothesized driver of growth in student social and emotional competence. To test this hypothesis would require measuring program implementation *and* the student outcomes the program is hypothesized to produce using assessments such as SELweb.

Limitations

Improvement science does not provide a perfect solution to SEL's problem of scale. First, improvement science is designed to solve local problems of practice that are largely defined in the field. In this chapter, we presume *a priori* that implementing SEL programs consistently and well is a relevant problem of practice across many educational settings. In addition, improvement science appears to require a level of expertise, intellectual

dexterity, and institutional culture that may be difficult to implement in practice without first engaging in significant systems change initiatives to accommodate the methods and habits of improvement science themselves. In light of these two limitations, it may be useful to consider models of continuous improvement that are more prescriptive and therefore may be easier for school districts to adopt and successfully implement. We consider one such model next.

Model 2: Positive Behavioral Interventions and Supports

What It Is

Positive behavioral interventions and supports, or PBIS, is a form of systems intervention that is adjacent to and complements SEL programs (www.pbis.org). PBIS is a system of daily practices and strategies aimed at setting clear positive behavioral expectations, reinforcing positive behaviors, and dealing constructively with behavioral infractions. The aim of PBIS is to improve student behavior, academic outcomes, and school climate, and reduce punitive disciplinary practices. From its inception, PBIS has integrated regular assessment data collection and review as a part of PBIS itself. Assessment elements that are routinely collected as part of PBIS include implementation data to understand how PBIS practices are being executed and student outcome data to understand the impact of those practices on behavioral and other outcomes. In addition, PBIS provides guidance about creating school-based teams to review and make decisions based on assessment data. Built into the PBIS practice model, therefore, are guidelines about what kinds of assessment data to collect to support practice, who to convene to collect and review those data, and how to structure teams and meetings to make data-based practice decisions.

One of the benefits of integrating assessment with PBIS is that we have a rich body of evidence about (1) supports required to obtain implementation fidelity (Barrett, Bradshaw, & Lewis-Palmer, 2008), (2) the impact of implementation on behavioral and other outcomes (Bradshaw, Mitchell, & Leaf, 2010), and (3) the impact and benefits of PBIS implementation on the organizational climate of schools (Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008). The published literature suggests that when PBIS is well implemented, it results in improved student behavior and reductions in disciplinary referrals (Bradshaw et al., 2010), improved climate (Bradshaw et al., 2008), and sometimes improved academic outcomes (Bradshaw et al., 2010). Those conclusions are helpful to decision makers who are considering adopting PBIS—they can weigh high-quality evidence of its impact in deciding whether it is right for them. This body of evidence is equivalent to meta-analyses of SEL programs, and the studies they summarize.

By integrating assessment into practice, PBIS continues its commitment to rooting practices in data. Integrating assessment with PBIS practice drastically reduces the chances that PBIS will be conflated with ineffective practices, or that watered down PBIS will lead to the incorrect blanket conclusion that PBIS does not work. Deeply integrated intentional assessment and data review practices reduce the chances that PBIS will fall prey to the “fidget spinner effect.” This is because PBIS practices and outcomes are well-measured, so it is much less likely that an ineffective practice or poorly implemented PBIS will be confused with a potent form of PBIS. For example, if PBIS were implemented in an elementary school and the PBIS team, in reviewing implementation and student outcome data, learned that implementation was inconsistent and that student outcomes were not moving in the right direction, they would have a basis for supporting changes in adult practice, a way to measure progress as adults adopted new ways of doing PBIS, and a way to see whether student outcomes were moving in the right direction. In short, the PBIS team and their colleagues at this school would be able to tell how well they were “doing PBIS,” what difference it was making, and whether changes in adult behavior were in order.

In addition, the continual use of data means that the field—including educators and scientists—will have useful information about how much and how well PBIS is practiced and how that changes student outcomes. When implementation is low or inconsistent, or outcomes are not moving in the right direction, this provides opportunities for reflection to diagnose the problem and approach intervention in new ways designed to improve practices and their downstream benefits to students. Integrating data systems into PBIS has supported a healthy growth and evolution of the field.

What the Field of SEL Can Learn

PBIS as a field of practice has integrated the kind of data surveillance systems that prevention scientists and others have identified as crucial to bringing effective interventions to scale while maintaining fidelity and impact. Because the target of PBIS is behaviors that are adjacent to, and overlap, the kinds of outcomes SEL programs seek to influence, it may therefore serve as an important model for integrating SEL assessment and its use into the practice of SEL. PBIS therefore has several important guideposts for the field of SEL.

First, the PBIS practice model is clear about what kinds of assessment data to collect, in ways that can be helpful to the field of SEL. PBIS practices include assessing (1) implementation, which includes what and how well PBIS practices are being put into practice, (2) climate, which involves student and teacher perceptions of the conditions of learning such as sense of

safety and connectedness, and (3) student outcomes, including administrative data reflecting disciplinary practices such as office discipline referrals. Although the goals and tactics used in SEL programs are distinct from PBIS, as I have suggested earlier, these general forms of assessment data stand to support the field of SEL in the same manner and magnitude that they support the field of PBIS. PBIS therefore provides a model of the *kinds* of assessments that might be helpful. SELweb and the associated measures of climate and program implementation include these elements, with the goal of providing the kind of data educators can use to reflect on what they are doing, and how it is affecting changes in student competence and the school climate. In creating a system that integrates measures of implementation, student competence, and climate, we hope to support the field of SEL in the ways that PBIS's integrated assessment practice supports high-quality PBIS.

Second, PBIS provides a model of *how* assessment data can be used effectively to support practice. Specifically, PBIS has developed a specific model for data use called "team-initiated problem solving." This involves a standardized process of problem identification, goal clarification, solution generation and evaluation, implementation monitoring, and impact evaluation. This fits nicely into the response to intervention (RTI) approach to special education (www.pbis.org/pbis/tiered-framework). To support PBIS educators in implementing this system, free resources are available such as meeting agenda templates and other guidance (e.g., www.pbis.org/resource/tips-meeting-minutes-template). Like the PBIS assessment, the Behavioral Incident Report System (BIRS), described by Fox, Veguilla, and Hemmeter (Chapter 8, this volume), offers a system for early childhood settings for assessing and addressing student behavior problems in a continuous improvement framework.

The field of SEL, in contrast, does not have a model of data review to guide educators in using whatever SEL assessment data they collect to make decisions, although some efforts have been made to articulate such a model (McKown, 2019a). A key lesson from PBIS, therefore, is that, beyond the kinds of SEL assessment data that are collected, for educators to use SEL assessment data constructively, they will benefit from support to develop an effective process for reviewing, reflecting on, and making decisions based on assessment data. Educators who use our teacher implementation survey, SELweb, and the student climate survey largely follow the PBIS way. That is, they assess periodically, review the data in teams, identify points to celebrate and areas of concern (in adult practices, climate, and student competence), and make specific plans intended to build on strengths and address needs. After a period of implementation, they reassess to see how student competence and climate are changing. They may also reassess teacher implementation to see how things have changed. The field of SEL has much to do to develop a robust data-monitoring and

evaluation capacity, but such systems are emerging and available to educators who understand how assessment can support effective practices.

Third, the PBIS assessment practice model articulates specific assessment goals. That is, in the PBIS framework, the goal of assessment is to provide the data needed to continuously improve PBIS practices and the impact they are having on the climate and student behavior. This is distinct from other assessment goals such as assessing to determine school quality, diagnose students, or screen and identify students for extra support.² The major emphasis of assessment in the PBIS framework is therefore continuous practice improvement. Such a focus places assessment squarely in service to supporting high-quality practice and avoids the potential political pitfalls of using social and behavioral data for accountability purposes. Similarly, and consistent with the recommendations of recent reports (Assessment Work Group, 2019; National Practitioner Advisory Group, 2019), the field of SEL would do well to be clear about the assessment goals that will (and will not) best support consistent and high-quality program implementation. As states adopt standards, it seems likely that SEL assessment will be called on to meet accountability goals. However, in my view, and the view of the Assessment Work Group and National Practitioner Advisory Group, using SEL assessment for accountability purposes poses some risks. If funding and other high stakes are attached to test results, SEL practices and assessment results may become distorted and become a source of resentment and resistance. If, on the other hand, assessment is used to support educators and students to achieve their goals, without the threat of negative consequences, it seems more likely that assessment will be able to fulfill the goal of supporting consistent, high-quality program use.

Limitations

There is a key difference between PBIS and SEL. Specifically, PBIS is a flexible set of principles and practices that can be adapted to the realities of different schools and districts. In contrast, SEL includes a loose confederation, bound by evidence of efficacy and the goal of promoting student social and emotional competence, of programs that are provided by different companies, program providers, nonprofits, and universities. Although there is some overlap between programs, each is somewhat distinct in terms of instructional approach, method, content focus, and intended outcomes. Because of this, from one program to another, what constitutes high-quality implementation, and the expected impact on climate and specific student outcomes, may not be the same. If the field of SEL were to adopt an

²Note, however, that one component of PBIS practice does involve screening and identification of students who need extra support.

assessment practice model akin to that of PBIS, then an important issue to resolve would be the extent to which a single assessment system (combining assessment tools and data use practices) could support the diversity of program types. At one extreme lies a single program-agnostic assessment system designed to measure implementation, student competence, and climate regardless of the specific SEL program in use. At the other extreme, each program would need a specific and distinct set of assessments to measure implementation, competence, and climate. Given the large number of SEL programs in the field, it is important to work out how best to develop, integrate, or adopt assessments and data use practices to support the field.

• A Vision for Assessment Integrated with SEL

A key premise of this chapter is that high-quality assessment is necessary for evidence-based SEL programs to successfully go to scale. In this case, success means that massively adopted SEL programs continue to measurably benefit student outcomes. Improvement scientists recognize the importance of assessment and evaluation systems to monitoring the implementation and impact of evidence-based programs as they go to scale. Improvement scientists recognize the importance of assessment in monitoring the implementation and impact of systems change efforts in schools and other systems. PBIS recognizes the importance of assessment in monitoring the implementation and impact of PBIS. In both cases, the purpose of assessment is to provide feedback to decision makers, including teachers, so they know what is happening and how much difference it is making. In these instances, the point of assessment is to provide data that decision makers can use to make decisions about how to modify and improve practice and the outcomes it is intended to produce.

Building on the good work that has been done to develop social and emotional competence and climate assessments, what would an assessment system look like that can support SEL at scale? What should it measure? How should educators use the assessment data? What purpose should SEL assessment serve that is valuable to educators and the field? We have described our assessments, and they serve as one example of what such a system might look like. More specifically, next, I develop a vision for the elements of an SEL assessment system that could serve to support the health of the SEL enterprise.

Data Elements

As I have suggested previously, for SEL assessment to support the field of SEL as it “goes big,” two broad kinds of assessment should be assessed in

an integrative way, much as it is in PBIS. First, SEL assessment needs to include the measurement of implementation, broadly defined as the extent and quality with which SEL programs are being executed in classrooms and schools. Second, SEL assessment needs to include the measurement of outcomes that SEL programs are intended to produce, including improvements in student social and emotional competence and improvements in school climate.

Intended Uses

What should the goal of SEL assessment be? How should assessment data be used? How should they not be used? The goal of SEL assessment is to provide information to decision makers—from superintendents to teachers—about SEL programs and key outcomes so that they can make decisions about how to support consistent, high-quality programs, and the positive outcomes those programs are intended to nurture. A superintendent or principal, for example, might use implementation data to identify schools or classrooms with low-frequency implementation and then guide coaching resources to those settings to support greater implementation. Pairing implementation and outcome data can provide important information about the extent to which engaging in SEL programs is associated with improvements in student competencies and climate in ways that suggest the program is working. The broad goal of SEL assessment is therefore continuous improvement—the use of data gathered at strategic points in the year to guide consistent, high-quality SEL program implementation to produce positive student outcomes. The goal of SEL assessment is *not* to measure school quality, to evaluate teacher performance, or to screen, diagnose, or label students.

Technical Requirements

What technical properties should these assessments have? It is beyond the scope of this chapter to offer an in-depth treatment of psychometrics. However, it is important that assessments focused either on implementation or outcomes should have adequate technical properties where *adequate* means that the evidence of the assessments' score reliability, internal structure, association with other variables, and the like support their intended uses (Kane, 2013; McKown, 2019a). Imagine, for example, that a district opts to measure SEL implementation through a quarterly teacher survey asking teachers to report how many SEL lessons they taught, and how well the lessons went. To be confident in such an assessment, it would be helpful to have evidence that scores on this assessment are associated with other implementation assessments such as direct observation. Alternatively, it would help to know how strongly scores on the implementation assessment

are associated with student outcomes such that higher teacher-reported implementation is associated with greater improvement in student social and emotional competencies.

Assessment Requirements

Implementing any initiative at scale is hard. That includes SEL assessment. Several things need to be true for SEL assessment to stand a reasonable chance of being used in the field. Assessments must be usable and feasible to administer. *Usable* means that with minimal or no training, an educator can administer an assessment and produce usable data in an informative reporting format. For example, a Web-based assessment is most usable if it offers single sign-on and is integrated with a district's data-management system. Scoring and reporting is most usable if it is automated and does not require tedious hand scoring. *Feasible* means that administering the assessment and using the assessment data can be accomplished within the resource constraints of the school. A particularly important resource is time. For example, an assessment should consume as little instructional or planning time as possible. Similarly, data use meetings, to the extent possible, should be incorporated into existing standing meetings so that no additional time needs to be "found."

Practice Model

Beyond these minimal, though critical, practice requirements, what elements of an assessment practice model should any SEL continuous improvement system incorporate? This may differ somewhat by what is being assessed. Ideally, implementation, for example, will be assessed frequently enough to guide resources to support consistent and high-quality SEL programs. Too long an interval between assessments allows variation in consistency and quality across classrooms to persist. Too short an interval taxes the system. Assessment of outcomes might be done in the fall and again after a period of instruction. Armed with assessment data on student competence, educators can decide what social and emotional competencies to emphasize; with assessment data on climate, educators can decide what adult practices they might deploy to improve upon the conditions of learning. Initial assessment therefore serves a formative purpose, guiding resources and instruction to build on strengths and address needs. Assessing competence and climate after a period of instruction provides summative information about student progress in response to instruction. Interim assessments may serve as a mid-point check to see if things are moving in the desired direction and to make adjustments as needed based on assessment findings.

Program-Agnostic or Program-Specific?

One of the questions raised earlier in this chapter concerned the extent to which a single set of SEL assessments can provide equally useful and relevant data regardless of the specific SEL program under consideration. Some might argue that each program is sufficiently unique that the way implementation is measured should be tailored to each program. Similarly, the intended student outcome or dimension of school climate might vary some from one SEL program to another, and so the outcome measures need to be specific to each program. The unfortunate consequence of this line of reasoning is that it would require a vast array of assessment tools that vary slightly in emphasis. Developing technically sound assessments is a costly proposition. As a result, creating separate assessments for each program will mean that inadequate resources may be invested in each, resulting in variable quality at best. Even if quality were not an issue, separate assessments for each program will result in difficulty comparing implementation quality and outcomes from one program to another. Furthermore, SEL program developers generally do not offer assessments along with their curricular or programmatic resources. Indeed, they may have a perverse incentive not to do so, as assessment might demonstrate that their program is not effective, placing them at a competitive disadvantage. As a result, it seems highly unlikely that program developers will be in a position to develop the kinds of high-quality assessments the field requires.

For many reasons, then, it seems desirable that the field develop a flexible suite of program-agnostic SEL assessments akin to those integrated into PBIS. Such a system would offer usable and feasible assessments focused on implementation, climate, and competence that can be used in conjunction with any SEL program. Ideally, such a system would offer flexibility within structure, so that district decision makers might customize, to a degree, how and what they assess, while maintaining core assessment elements in every setting. Many forces would have to converge to realize such a vision. Nevertheless, were the policy context to support this general proposition, and sufficient funding available to support its development and implementation, this is an achievable vision. In fact, it would not require a whole-cloth invention of assessments, but could conceivably involve integrating or adapting the best available assessment tools into a single delivery system optimized for usability and feasibility.

Data Use Practices

Assessment tools themselves cannot benefit instruction or student outcomes unless educators review and reflect on assessment results, and make decisions about resources and instruction based on what they learn. In other

words, data use practices are an essential element of assessment to support SEL program use. It is beyond the scope of this chapter to describe effective data use practices in depth. However, key elements include (1) leadership support for data use, (2) a shared understanding of what decisions will be made based on the assessment results, (3) a shared understanding of the meaning of assessment scores and reports, (4) access by decision makers to the assessment data, (5) opportunities to reflect on the meaning of the data independently, (6) regular time to meet for data review, and (7) structured group discussion of assessment results that includes reviewing the facts, interpreting their meaning, and developing a plan of action (McKown, 2019a). Data use in schools varies tremendously from district to district, and depends on a number of factors, including access to data, leadership, norms, routines, and time (Coburn & Turner, 2011; McKown, 2019b). To the extent that integrated SEL assessment systems can also provide guidance for data use, much in the way PBIS does, this will lower barriers to effective data use.

• Pipe Dream or Possible?

One might reasonably ask whether the vision of an SEL assessment system designed to support ongoing SEL programs and impact is attainable. After all, assessment development is costly, and adding yet another administrative and assessment burden to school systems that are already brimming with regulatory obligations seems like a hard sell. The easiest path—in fact, the default position of the field—is that we should focus on programs, and that assessment is a luxury item to be used as time, interest, and resources permit. In short, the default position is to skip assessment altogether and operate under the assumption that if a program is evidence-based, it will continue to produce positive outcomes at scale.

The default position is risky. If educators do not assess the use of programs, and the outcomes they are intended to produce, the field will be susceptible to the “fidget spinner effect” described earlier in the chapter—all manner of ineffective activities will be ready to co-opt the label “SEL” while it is a hot topic and word will eventually spread (incorrectly) that SEL includes weak and faddish programs and/or the little evidence that is collected will demonstrate that it does not work. And that will mark the beginning of the end of the field of SEL. So the question ought not to be, can we afford to develop and integrate SEL assessment with SEL programs? It ought to be, can we afford not to do so? I believe the answer is that in the short run, we can; but soon, the field will pay a heavy price and be overrun by some new educational fad.

It is heartening that there are examples of educators integrating SEL assessment in programming to support consistent and high-quality program use (see McKown & Herman, 2020). The Wisconsin Department of Public Instruction, for example, has adopted social and emotional standards and is providing districts with resources to identify programs and competence assessments targeting the competencies in its standards. The CORE districts in California use social and emotional assessments—in this case self-reported student competence—as part of ongoing school improvement efforts. And forward-thinking school districts are using our direct assessment of student competencies, often in conjunction with implementation and climate assessment, to shine a light on the use of SEL programs, and the outcomes those programs are intended to produce. In those districts, educators benchmark competence and climate early to guide practice; they then monitor implementation and provide support as needed; and they then reassess competence and climate to measure change.

There are important questions about the level at which SEL assessment should be adopted and used to guide practice. At one extreme, although it is unlikely, one could imagine a national SEL assessment program. At another, one could imagine teachers deciding whether and how to assess SEL. It is difficult to imagine a national SEL assessment program that would be sensitive to local district needs. At the other extreme, it is difficult to imagine assessment data would be broadly useful if selected by individual teachers. States may play a role in guiding districts by, for example, providing guidance on what to assess and a range of options for how to assess those things. In our work, we have found that districts vary in their needs and their readiness to adopt SEL programs and SEL assessment to support program use. It therefore seems sensible to focus on statewide guidance for local decision making about SEL assessments and their uses (McKown & Herman, 2020). Ultimately, this is an important issue for the field and for policymakers to take up.

The assessment practice examples cited above speak to an appetite for assessment, and a broadly shared instinct that SEL assessment can support program use. However, there are some natural limits on what is possible in the current context. First, the field of SEL assessment, like the field of SEL programs, is a loose confederation of organizations each offering somewhat different assessment systems, each suited to achieving somewhat distinct assessment goals. Second, some assessment providers focus on student competence and some focus on climate, but very few focus on both, and very few also offer measures of implementation. Third, and perhaps obviously from the foregoing, few assessment systems measure program implementation, student competence, and climate in the ways that are needed to support consistently high-quality SEL programs at scale.

If there is broad agreement that high-quality assessment data are important for the field's health, longevity, and impact, the big question is, what would need to change to facilitate the integration of usable, feasible, program-agnostic assessments that measure implementation and outcomes? McKown and Herman (2020) described conditions that can support the use of SEL assessment to improve program use. Key among their recommendations were policies that incentivize the use of SEL assessment, that guide the manner of their use, and that support educators in using SEL assessment data to guide the consistent and high-quality execution of evidence-based programs. In answer to the question of whether the vision of integrating SEL assessment into programs in the way that PBIS has is a pipe dream or is possible, the answer depends, in this as in many things, on whether educators and policymakers can muster the will and foresight to invest in and support an unglamorous but foundational support for maintaining quality at scale. Perhaps visions of fidget spinners shimmering in classrooms across the heartland will motivate the field toward a commitment to systematic and ongoing assessment, so that hard-won evidence-based programs are not undermined by the encroachment of ineffective programs, be they fidget spinners or otherwise.

REFERENCES

- Assessment Work Group. (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future*. Collaborative for Academic Social and Emotional Learning.
- Atwell, M., & Bridgeland, J. (2019). *Ready to lead: A 2019 update of principals' perspectives on how social and emotional learning can prepare children and transform schools*. Civic Enterprises.
- Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Interventions*, 10, 105–114.
- Bradshaw, C. P., Koth, C. W., Bevans, K. B., Jalongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly*, 23, 462–473.
- Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, 12, 133–148.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2017). *Learning to improve: How America's schools can get better at getting better*. Harvard University Press.
- Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement*, 9, 173–206.
- 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, 405–432.
- Dusenbury, L., Yoder, N., Dermody, C., & Weissberg, R. P. (March, 2020). *An examination of K–12 SEL learning competencies/standards in 18 states*. Collaborative for Academic Social and Emotional Learning. <https://casel.org/wp-content/uploads/2020/03/CSI-Frameworks.pdf>.
- Fagan, A. A., Bumbarger, B. K., Barth, R. K., Bradshaw, C. P., Cooper, B. R., Supplee, L. H., & Walker, D. K. (2019). Scaling up evidence-based interventions in U.S. public systems to prevent behavioral health problems: Challenges and opportunities. *Prevention Science*, 20, 1147–1168.
- Hamilton, L. S., Doss, C. J., & Steiner, E. D. (2019). *Teacher and principal perspectives on social and emotional learning in America's schools: Findings from the American educator panels*. RAND Corporation.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625–638.
- Jones, S., Bailey, R., Brush, K., & Kahn, J. (2017). *Kernels of practice for SEL: Low-cost, low-burden strategies*. Harvard University EASEL Lab.
- Kane, M. T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50, 1–73.
- Mahoney, J. L., Durlak, J. R., & Weissberg, R. P. (2019). An update on social and emotional learning outcome research. *Kappan*, 100, 18–23.
- McKown, C. (2017). Social-emotional assessment, performance, and standards. *The Future of Children*, 27, 157–178.
- McKown, C. (2019a). *Assessing students' social and emotional learning: A guide to meaningful measurement*. Norton.
- McKown, C. (2019b). Challenges and opportunities in the applied assessment of student social and emotional learning. *Educational Psychologist*, 54, 205–221.
- McKown, C., Allen, A. A., Russo-Ponsaran, N. M., & Johnson, J. K. (2013). Direct assessment of children's social-emotional comprehension. *Psychological Assessment*, 25, 1154–1166.
- McKown, C., & Herman, B. A. (2020). *SEL assessment to support positive social and emotional development*. The Pennsylvania State University.
- National Practitioner Advisory Group. (2019). *Making SEL assessment work: Ten practitioner beliefs*. Collaborative for Academic Social and Emotional Learning. Available at <https://measuringcasel.org/wp-content/uploads/2019/09/NPAG.pdf>.
- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83, 357–385.