Social-Emotional Assessment, Performance, and Standards

Clark McKown

Summary

In the push to boost young people's social and emotional learning (SEL), assessment has lagged behind policy and practice. We have few usable, feasible, and scalable tools to assess children's SEL. And without good assessments, teachers, administrators, parents, and policymakers can't get the data they need to make informed decisions about SEL.

Some existing SEL assessments, writes Clark McKown, are appropriate for some purposes, such as keeping teachers abreast of their students' progress or evaluating SEL interventions. But too few high-quality SEL assessments are able to serve a growing range of purposes—from formative assessment to accountability, and from prekindergarten through high school.

McKown recommends proceeding along two paths. First, he writes, educators should become familiar with existing SEL assessments so that they can learn their appropriate uses and limits in a low-stakes context. At the same, we need to invest money and talent to create assessment systems that can be used to meet important assessment goals at all grade levels.

McKown walks us through definitions of SEL, identifying three broad areas of SEL skills—thinking, behavior, and self-control. Each area encompasses skills that are associated with important life and academic outcomes, that are feasible to assess, and that can be influenced by children's experiences. Such *meaningful, measurable,* and *malleable* skills, McKown argues, should form the basis of SEL assessments.

The next generation of SEL assessments should follow six principles, he concludes. First, assessments should meet the highest ethical and scientific standards. Second, developers should design SEL assessment systems specifically for educational use. Third, assessments should measure dimensions of SEL that span the three categories of thinking, behavioral, and self-control skills. Fourth, assessment methods should be matched to what's being measured. Fifth, assessments should be developmentally appropriate—in other words, children of different ages will need different sorts of assessments. Last, to discourage inappropriate uses, developers should clearly specify the intended purpose of any SEL assessment system, beginning from the design stage.

www.futureofchildren.org

Clark McKown is an associate professor in the Department of Behavioral Sciences at Rush University Medical Center. This article was supported by grants from the US Department of Education, Institute of Education Sciences to Rush University Medical Center (#R305A110143, #R305A140562). McKown is also the founder and president of xSEL Labs, which markets the SELweb assessment tool.

Pat Kyllonen of Educational Testing Service reviewed and critiqued a draft of this article.

Social and emotional learning, or SEL, includes a broad range of mental, behavioral, and self-control skills that people use in social interactions to achieve social goals. Although scholars haven't reached consensus on its definition, SEL includes skills such as the ability to infer others' thoughts and feelings (thinking skills), the ability to initiate a positive interaction (behavioral skills), and the ability to stay calm when upset (self-control skills). Labeled variously as "soft" or "noncognitive" skills, SEL skills are highly consequential. Decades' worth of research has consistently found that the better developed their SEL skills, the better children do in school and life.[1]

Parents, educators, and policymakers increasingly recognize the importance of SEL. In the past three decades, prevention scientists and others have developed and rigorously evaluated a number of comprehensive, evidence-based SEL programs. These programs are widely used: In a 2015 nationwide survey of 562 teachers and administrators, 59 percent of respondents reported using a program called School Wide Positive Behavioral Intervention and Supports (SWPBIS), and 32 percent of respondents reported using an SEL program such as PATHS® or Second Step.[2] Furthermore, a growing number of states now include SEL in their educational standards.[3]

Purposes of SEL Assessment and Lack of Appropriate Tools

Although policy and practice are moving forward, one area lags. We have few usable, feasible, and scalable tools for educators to assess children's SEL, creating a conundrum for policymakers and practitioners. Just like good academic assessment, good SEL assessment could help educators achieve many goals. It could be used to determine children's strengths and needs, and guide decisions about curriculum and instruction; that's *formative assessment*. It could tell us whether SEL programs and practices work; that's *program evaluation*. It could be used to monitor students' social-emotional development in response to the introduction of interventions; that's *progress monitoring*. It could help determine whether children are meeting SEL standards; that's *standards-based assessment*. Finally, assessment could help decide whether students receive special services, and it can guide teacher, school, and district accountability; those are examples of high-stakes decision making based on SEL assessment data.

Here is the conundrum: Without good assessment, it's difficult to see how teachers, administrators, parents, and policymakers can get the data they need to make informed decisions as they seek to foster children's healthy social and emotional development. Without meaningful assessment data, decisions affecting children—from policy to instruction—are likely to be buffeted by the forces of fad and politics. For SEL policy and programs to be as effective as possible, we need to develop usable, scalable, and scientifically sound SEL assessment systems.

In addition, existing policy motivates practitioners to use SEL assessment for some purposes more than others. In particular, a growing number of states have incorporated SEL components into their learning standards, creating a powerful impetus for educators to select and develop curriculum materials and instructional

strategies to ensure that students meet those standards. At the federal level, the Every Student Succeeds Act, or ESSA, gives states flexibility to use nonacademic assessments of school environment and student outcomes for accountability.

In light of state standards and federal law, it seems likely that SEL assessments will be called upon to determine whether teachers, schools, districts, and states are successfully fostering social and emotional outcomes alongside academic ones. That's a problem because, broadly, no system of social-emotional assessment that I'm aware of has adequate technical properties to serve as part of a high-stakes accountability system. Current SEL assessments, many of which I describe below, are appropriate for formative assessment, program evaluation, and progress monitoring. They may also be appropriate for low-stakes measurement of progress toward state standards, by which I mean broad surveillance to determine whether schools, districts, and states are moving in the right direction, without high-stakes consequences attached. All of these purposes, and the assessment systems available to fulfill them, may put us on a path toward SEL assessment for high-stakes accountability. But prematurely adopting assessments ill-suited to accountability may inadvertently undercut advances in the field of SEL.

Thus, we see a mismatch between what's arguably the greatest demand for assessment—high-stakes accountability—and the appropriateness of existing assessment systems. This problem has no easy solution. However, two constructive parallel paths may help maximize benefit while mitigating risk. First, educators should become familiar with, adopt, and use existing, well-designed SEL assessments for appropriate purposes—formative assessment, progress monitoring, and program evaluation—so that they can learn their uses and limits in a low-stakes context. Second, a significant investment of money and talent will be needed to create assessment systems that can serve multiple ends, including ends such as high-stakes accountability for which existing assessments are inappropriate.

Because SEL assessment systems are underdeveloped, it's important that schools and districts undertake SEL assessment with clear goals and realistic expectations. Contrast fictitious districts A and B. Leaders of District A have decided to measure many dimensions of SEL and to determine how to use those measures afterward. Leaders of District B have decided to measure particular SEL skills exclusively to guide instructional planning. Because District A doesn't make clear how SEL assessment data will be interpreted and used, there's a strong possibility that the data could serve inappropriate purposes, such as evaluating teacher performance. And because District A isn't clear about goals, it's likely that it will expend considerable resources gathering data that aren't put to work to help teachers teach and children grow.

In contrast, everyone involved in District B knows the purpose of assessment and the uses of assessment data. Because the purpose is clear, the district can arrange focused and practical training in how to interpret and use the assessment data, increasing the odds that they will be used appropriately. Moreover, everyone involved in District B understands that a large number of decisions—school and teacher accountability, special education placement,

etc.— won't be guided by the data. Therefore, educators will be less anxious that data could be used against them. It's still possible that formative assessment data collected in District B could do unintended harm. But because the goals are clear, that's significantly less likely.

Neither researchers nor practitioners nor policymakers have come to a consensus about what SEL is.

Before practitioners, program evaluators, policymakers, and others can use SEL assessment for any purpose, we need to define SEL and identify which dimensions can and should be measured for what purposes. Practitioners should also consider what methods of assessment are best suited to measuring a targeted SEL skill.

What Is SEL?

To create SEL standards and assess progress toward those standards presupposes that we agree about what SEL is. Yet neither researchers nor practitioners nor policymakers have come to such a consensus. The Collaborative for Academic Social and Emotional Learning (CASEL) defines SEL broadly 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." The CASEL model names five categories of SEL skills: self-awareness, self-management, social awareness, relationship skills, and responsible decision making.[4] This widely cited model has influenced the content of state SEL standards.

Other models complement or compete with CASEL's. A report on "Foundations for Young Adult Success" from the University of Chicago describes "noncognitive" factors that include academic behaviors, academic perseverance, academic mindsets, learning strategies, and social skills.[5] Another report, by the National Academy of Sciences, argues that "21st century skills" include both intrapersonal or self-management skills and interpersonal or people skills.[6] Other scholars emphasize cognitive, emotional, and social/interpersonal skills, along with the school context that influences how those skills develop and the outcomes they produce.[7] Still others emphasize information processing, emotional processes, or argue that attitudes such as grit or growth mindsets are part of SEL.[8]

Each of these models has merit, and each of the skills, competencies, behaviors, and attitudes they describe is consequential. But to have competing models that claim to describe the same thing can cause problems. It interferes with communication (we use the same words to mean different things), impedes science (we can't accumulate knowledge on SEL if each researcher has a different definition), undermines practice (dissimilar programs with unequal effectiveness can be described with the same language), and confuses the public. In addition, when policymakers genuinely interested in fostering children's SEL turn to experts for guidance, they may get conflicting advice that could become codified into a crazy quilt of standards. Arguably, the assessment endeavor suffers

most from conflict over what SEL is, as definitional ambiguity makes it hard to translate good ideas into sound assessment practices. Vigorous efforts to create conceptually coherent and scientifically sound SEL assessments may help to create a common understanding of SEL.

Finding Common Ground for Policy, Practice, and Assessment

Despite their differences, all models of SEL share important commonalities. Most describe skills used in social interactions that vary across individuals, that are associated with important interpersonal and life outcomes, and that are malleable. In addition, all models of SEL encompass three broad categories: thinking skills, behavioral skills, and self-control skills. We can find common ground, therefore, by defining SEL as *the thinking, behavioral, and self-control skills that are applied in social interactions and that influence children's social and other life outcomes*. That definition is sufficiently specific to guide policymakers, practitioners, and assessment developers, but sufficiently flexible to spur continued innovation.

SEL thinking skills include the ability to encode, interpret, and reason about social and emotional information, as we do when we recognize others' emotions, take others' perspectives, or solve social problems.[9] SEL behavioral skills are actions people take during social interactions to achieve a social goal. Behavioral skills include positive actions that are associated with making and maintaining positive relationships, such as assertiveness, politeness, and turn-taking. But social behavior also includes negative actions that interfere with positive relationships, such as aggressiveness, impulsivity, and social withdrawal.[10] Self-control skills encompass the ability to modulate thoughts, feelings, and behavior to achieve a goal.[11] In this article, I focus on self-control as applied in social contexts. Some dimensions of self-control are mental, including the effortful control of attention and emotions.[12] Other dimensions of self-control are behavioral, such as refraining from impulsive behavior.

Precisely what thinking, behavioral, or self-control skills make up SEL is an open question. I next describe SEL skills that are *meaningful, measurable,* and *malleable*.[13] To be meaningful, SEL skills must be associated with important life and academic outcomes, and included in SEL policies and programs. To be measurable, SEL skills must be feasible to assess. *On-task behavior*, for example, is a measurable skill, while *virtue* is a construct that's more difficult to measure. To be malleable, SEL skills must be influenced by experience, as demonstrated either by observational research establishing a relationship between experiences and skills or by studies demonstrating that a particular intervention can influence a targeted skill. Within each of the three areas of SEL—thinking, behavior, and self-control—we can identify skills that are meaningful, measurable, and malleable.

Meaningful, Measurable, and Malleable Dimensions of SEL

Meaningful Thinking Skills

Several SEL thinking skills are meaningfully related to important outcomes and have been incorporated in state standards. For example, children with a better-developed *ability to recognize emotions in others* do better on a range of important outcomes. Research has shown that, for example,

preschoolers' knowledge of emotions predicts concurrent and later social competence and academic success.[14] This association persists into the elementary grades; a review of 14 studies and found that in first through sixth grade, children who were better at reading emotions from facial expression, tone of voice, and posture also had better-developed reading and math skills. The same review also found that being able to recognize emotions was positively associated with self-control, self-esteem, and peer acceptance.[15]

Perspective-taking—defined as the ability to infer others' beliefs, thoughts, and desires—is also meaningful. For example, several investigators have found that preschoolers' understanding of others' beliefs and perspectives is associated with later academic skills.[16] But perspective-taking's benefits extend well beyond academic outcomes. Research has shown that children who are better at inferring others' beliefs are more prosocial, less aggressive, less withdrawn, and more accepted by peers.[17]

Social problem-solving involves understanding interpersonal conflict, developing social goals, and generating ideas about how to resolve those conflicts. Among school-age children, social problem-solving is associated with academic functioning.[18] In addition, children who are better at solving social problems are less aggressive and more frequently engage in socially positive behavior.[19]

Together, emotion recognition, perspective-taking, and problem-solving are more strongly associated with positive academic and social outcomes than any one of them is in isolation. For example, when we examined both typically developing and clinic-referred children from four to 17 years old, our group of researchers found that a composite score reflecting emotion recognition, perspective-taking, and social problem-solving together was robustly associated with positive social behavior as reported by parents and teachers. The magnitude of that association was greater than the magnitude of associations between any individual skill and behavioral outcomes. Our finding suggests that we should be assessing multiple dimensions of SEL thinking skills.[20]

SEL thinking skills change with age, in both quantity and quality.

The dimensions of SEL that I've discussed are reflected in the Illinois state SEL standards, the first comprehensive preschool through high school state SEL standards in the nation. For example, the standards declare that upper elementary children should be able to "identify verbal, physical, and situational cues that indicate how others may feel" and should be able to "describe the expressed feelings … of others" (emotion recognition); "describe the … perspectives of others" (perspective-taking); and "manage and resolve interpersonal conflicts in constructive ways," "apply decision-making skills to deal responsibly with academic and social situations," and "identify the steps of systematic decision-making" (problem-solving).[21]

Developmental Considerations

SEL thinking skills change with age, in both quantity (overall skill level) and quality (the

kinds of social-emotional phenomena that children can understand). For example, we know that between childhood and adulthood people's ability to recognize basic emotions from facial expression improves significantly.[22] We also know that new kinds of understanding of others' emotions develop in late elementary school, when, for example, children become capable of understanding that people can feel mixed emotions and that the morality of actions is associated with awareness of complex emotions such as pride and guilt.[23]

Similarly, children's understanding of others' perspectives develops with age, and these changes, like changes in emotion recognition, are both quantitative and qualitative. Between the ages of three and six, children develop a more and more advanced understanding of others' beliefs and desires;between eight and 12, we see a dramatic increase in children's ability to infer others' beliefs in real-world contexts.[24] Furthermore, children come to understand the relationship between thoughts, emotions, and behavior in themselves and others.[25]

Fewer researchers have examined age-related changes in social problem-solving, but we do have evidence that social problem-solving skills improve with age; that what constitutes a competent response changes somewhat with age—for example, asking for adult help is considered more competent when children are younger than when they're older; and that nevertheless, the components of social problem-solving don't change throughout the lifespan.[26]

Measurable Thinking Skills

Researchers have developed direct assessments for emotion recognition, perspective-taking, and social problem-solving.[27] As with all assessments, these tests have strengths and weaknesses. Most assess a particular dimension of social-emotional comprehension. Few are suitable for mass administration. We have no usable, feasible, and scientifically sound system that can be administered to groups of children to assess social-emotional comprehension and execution in the upper elementary grades.

To investigate the feasibility and promise of measures of social-emotional thinking, my colleagues and I collected data from 186 general education students and 118 clinic-referred children ages six to 14, using direct assessments that had been designed for research purposes. We found that:

- emotion recognition, perspective-taking, and social problem-solving can be reliably assessed;
- these three constructs are partially independent components of a higher-order global social-emotional comprehension construct;
- individual children's social-emotional comprehension varies considerably;
- general-education students perform better than clinic-referred children on direct assessments of social-emotional comprehension, and;
- better social-emotional comprehension is associated with more frequent socially competent behavior and less frequent socially aversive behavior, such as aggression, impulsivity, norm-violating aberrant behavior, and social withdrawal.[28]

Malleable Thinking Skills

Several lines of research suggest that SEL thinking skills are malleable. Evidence-based SEL programs include a range of curricula and instructional strategies designed to promote social-emotional comprehension and execution among all students. Children who participate in well-implemented, evidence-based SEL programs do better on measures of social, behavioral, and academic outcomes. A 2011 meta-analysis summarized the impact of 213 school-based universal SEL programs that included 270,034 students. It found that when the programs were implemented well, about 67 percent of children improved in their thinking skills, compared with about 34 percent of children who didn't participate in the programs.[29]

These studies suggest that SEL thinking skills are malleable. But they focus on programs that have many components and they measure multiple outcomes, leaving open the question of which skills are most malleable and what interventions are most effective for what skills. Some research suggests that targeted interventions can influence specific SEL thinking skills. For example, our group and others have found that when facial emotion-recognition training technology is paired with individual coaching, children can learn facial emotion-recognition skills. Similarly, interventions for high-functioning children on the autism spectrum have improved their perspective-taking skills, and interventions to teach social problem-solving skills have been effective for children with aggressive behavior.[30] Together, these studies suggest that specific SEL thinking skills are malleable.

Meaningful Behaviors

SEL encompasses both socially skilled behaviors, characterized by positive interactions that enhance relationships, and socially aversive behaviors, characterized by negative interactions that detract from relationships.[31] Behavioral skills are associated with academic and other important outcomes. In one study, for example, first- through sixth-grade students' interpersonal skills, as reported by their teachers, were positively associated with their standardized test scores.[32] In a sample of 423 sixth- and seventh-graders, more positive social behavior was associated with better grades and test scores.[33] Longitudinal studies—that is, studies that followed students over time—have found that positive social behavior in third grade is associated with greater academic achievement in eighth grade and that children who exhibit prosocial behavior in kindergarten are likely to attain more years of schooling.[34] In contrast, socially aversive behaviors are associated with poor academic outcomes. For example, aggressive behavior in kindergarten predicts lower scores on standardized literacy and math tests in later grades.[35] Among school-aged children, hyperactivity and impulsivity are also associated with poor academic outcomes.[36]

Behavioral skills are also associated with nonacademic outcomes. For example, elementary school–age children who rarely exhibit socially skilled behavior and more frequently exhibit socially aversive behaviors are more likely to be socially rejected; in turn peer rejection puts children at risk for maladaptive behavior and poor mental health.[37] Similarly, prosocial skills in kindergarten are associated with greater adult employment and a lower likelihood of

using public assistance, exhibiting criminal behavior, or suffering from mental illness.[38]

Social behaviors are integral to some state standards. For example, the Illinois standards say that children should learn to "identify and manage [their] … behavior," "demonstrate ways to express emotions in a socially acceptable manner," "manage and resolve interpersonal conflicts in constructive ways," "apply constructive approaches in resolving conflicts," and "use communication and social skills to interact effectively with others." [39]

Developmental Considerations

As we've seen, the quantity and quality of thinking skills change as children grow older. But socially skilled and socially aversive behaviors remain somewhat more stable across the elementary grades. Impulsive and aggressive behaviors do typically decline from early childhood through adolescence, and children learn about and can express increasingly complex positive social behaviors as they grow older. In general, however, similar positive and negative behavioral skills are important at all ages—a fact that's reflected in the construction of many widely used behavior rating scales. For example, the Social Skills Improvement System, or SSIS, has a single form for children from five to 12 years old, and the averages and distributions of the rating scale's scores don't change dramatically in that age range. [40]

Measurable Behaviors

Many rating scales measure important dimensions of social behavior. Whatever their focus, they ask raters (usually teachers) to assess the frequency of various behaviors.

Like SEL thinking skills, positive and aversive social behaviors appear to be malleable.

The resulting scores tell us how the frequency of a child's behaviors compares to the frequency of those behaviors in a typical sample. Many such scales are well-suited to measuring behaviors that either support or interfere with positive social relationships because behavior, unlike social-emotional comprehension, can be directly observed. For example, the SSIS assesses several dimensions of socially skilled and aversive behaviors. The Devereux Student Strengths Assessment focuses specifically on social-emotional learning skills, such as relationship skills and goal-directed behavior.[41] Using both teacher ratings and children's own reports, the Academic Competence Evaluation Scales measures both academic competence and the socially skilled behaviors associated with it.[42]

Malleable Behaviors

Like SEL thinking skills, positive and aversive social behaviors appear to be malleable. We have evidence from three kinds of interventions—those designed to nurture children's social-emotional skills, those designed to create social norms that influence behavior, and those that use instructional strategies to reduce problem behavior.

Programs that nurture social-emotional skills. Meta-analyses suggest that school-based SEL programs produce significant and meaningful behavioral benefits. One meta-analysis of 213 universal school-based SEL programs found that about 57 percent of children who participated in well-implemented, evidence-based SEL programs showed improvement on measures of behavioral outcomes, compared with about 43 percent of children who didn't participate in such programs.[43] Programs for children with psychological and behavioral problems showed even more dramatic benefits: In a meta-analysis of 130 indicated preventive interventions, about 63 percent of children improved, compared with about 36 percent of children in control groups.[44]

Programs to create social norms. SWPBIS encompasses a universal framework for behavior management that applies broad and flexible principles, rather than prescribed programs. In SWPBIS schools, educators set and teach positive behavioral expectations, collect and review data on student behavior, and use various strategies to encourage desired behaviors and discourage undesired behaviors. In a randomized field trial involving 37 ethnically and socioeconomically diverse schools with more than 12,000 students, children in SWPBIS schools displayed more positive behaviors and fewer problem behaviors than children in schools that didn't use SWPBIS.[45] In another study, students with greater teacher-reported concentration problems, disruptive behavior, and emotion dysregulation, and less frequent positive behavior, showed the greatest increases in positive behavior and the greatest decreases in disruptive behavior when exposed to SWPBIS.[46] Thus, SWPBIS provides further evidence that social behavior is malleable.

Instructional strategies. Instructional strategies can increase positive social behaviors and reduce problem behavior. Take, for example, the Good Behavior Game, in which children are assigned to groups and given points for targeted misbehaviors. The team with the fewest points wins a prize after a specified number of rounds. (It can also work the other way around: teams get points for positive behavior.) Many studies have shown that the Good Behavior Game significantly reduces problem behaviors.[47] The game is relatively easy to implement well, is widely accepted by teachers, and can be incorporated into regular classroom curricula.

Meaningful Aspects of Self-Control

It's beyond the scope of this article to review the complex scholarship about self-control and the debates about how to define, measure, and assess it. Rather, I'll examine three interrelated and commonly studied dimensions of self-control that are known to be associated with social and academic outcomes: delayed gratification, frustration tolerance, and behavioral impulse control.

A recent review of research found consistent evidence that cognitive, social, and emotional dimensions of self-control are all associated with young children's readiness to enter school.[48] Moreover, self-control remains important throughout elementary school. For example, in a study of an ethnically diverse sample of six- to 10-year-olds that relied on reports from teachers, effortful control was positively related to academic skills.[49] And in two samples totaling 304 middle school students, a measure of self-control that integrated parents', teachers' and children's own reports of their behavior with performance

on a delay of gratification task were better than IQ at predicting eighth-graders' academic outcomes.[50]

Indeed, childhood self-control is associated with wellbeing throughout the lifespan. Analyzing data from the Dunedin Multidisciplinary Health and Development Study—a longitudinal study of more than 1,000 participants who were followed from birth to adulthood—one set of researchers found that after controlling for initial socioeconomic status, a composite measure of researcher-observed and parent- and teacher-reported impulsivity in childhood was strongly associated with adult outcomes as wide-ranging as physical health, substance use, income, socioeconomic status, single parenthood, and criminality.[51]

Self-control is also incorporated in the Illinois SEL standards, which state that children should be able to "identify and manage … emotions and behavior" and "describe and demonstrate ways to express emotions in a socially acceptable manner."[52]

Developmental Considerations

Self-control measured in childhood is strongly associated with both concurrent and later outcomes. As with other dimensions of social-emotional learning, children's self-control changes with age. In early childhood, behavioral impulsivity is sufficiently typical that behavioral performance tasks, such as the famous marshmallow task I discuss in the next section, are meaningful indicators of self-control. In early elementary school, however, behavioral self-control becomes better developed and the frequency of impulsive behavior declines. In fourth to sixth grades, children can use attentional, cognitive, and behavioral strategies to control their behavior.[53] Because self-control changes with age, the means of measuring it must also change with age.

Measurable Aspects of Self-Control

How is self-control best measured? In preschool, simple behavioral-challenge tasks measure delay of gratification. For example, in the marshmallow task, children must choose between an immediate reward of a marshmallow and a larger but delayed reward of several marshmallows.[54] More recently, researchers developed the Preschool Self Regulation Assessment, which uses a series of simple performance tasks, from holding a piece of candy on the tongue to walking slowly on a line, to measure different aspects of self-control.[55] Scores from the assessment are reliable (consistent across tasks and time) and valid (associated with other measures of self-control), and are associated with social competence and school readiness.[56]

Beyond preschool, various direct assessments have been developed to measure mental aspects of self-control. Our team developed two web-based direct assessments for children in kindergarten through third grades to measure self-control. The first was a choice-delay task in which children chose between lower-scoring but fast responses and higher-scoring but slow and tedious responses.[57] The second was a frustration tolerance task in which children were given a certain amount of time to solve a problem; to induce frustration, the task was programmed so that several items stuck, as if the computer had frozen.[58] Both tasks yielded reliable scores that were associated with other social-emotional thinking skills and functional outcomes.

Other strategies to directly assess aspects of self-control have shown evidence of feasibility, including:

- asking children to follow rules that require them to disregard their natural inclinations, such as directing them to press the right side of a screen when something appears on the left side;[59]
- using a computerized game called the Iowa Gambling Task to measure the tendency to select smaller consistent rewards over large but risky rewards, and;[60]
- asking children to choose between a series of smaller but more immediate rewards and larger but delayed rewards.[61]

In the elementary grades, teaching children mindfulness shows promise to improve both self-control and other dimensions of wellness.

Malleable Aspects of Self-Control

Some evidence suggests that self-control is malleable. For example, when the Chicago School Readiness Project (CSRP)—an intervention to train teachers in behavior management strategies that can foster student self-regulation—was tested in a randomized field trial in Head Start preschools, children whose teachers received the training showed both greater self-control and stronger early literacy and mathematics skills. Furthermore, the investigators found some evidence that improved self-regulation was the mechanism through which the CSRP intervention improved early academic skills.[62]

In the elementary grades, teaching children mindfulness—the ability to focus attention on present experience without judgment—shows promise to improve both self-control and other dimensions of wellness. One randomized study of a brief mindfulness intervention among fourth- and fifth-graders found that students who learned mindfulness strategies improved their cognitive control and showed fewer physiological signs of stress. Moreover, the children who participated in the intervention were better liked by their peers, who said that the participants exhibited positive behavior more often.[63]

The Right Tool: Matching Method to What's Measured

The match, or lack thereof, between the measurement method and the dimension of SEL being measured is a critical and underappreciated consideration. *Method* means the procedure we use to sample behaviors that are hypothesized to reflect an underlying skill; they include self-report questionnaires, peer nominations or ratings, observation, teacher ratings, a hybrid of observation and teacher ratings called direct behavior ratings, and direct assessments, in which children demonstrate skills by solving challenging problems.[64] No single method measures every dimension of SEL well, and each is better suited to measuring some things than others.

Thinking skills, behavioral skills, and self-control skills are each best measured in different ways. For example, although observers and raters can make educated

guesses about children's thinking skills, these skills exist in a child's mind and can't be directly observed. A skill such as reading others' facial expressions is an unobservable mental event. To assess it through observation requires a large inferential leap from observable behavior. The same is true for perspective-taking and problem-solving skills.

So although teachers could rate such skills or children could rate themselves, direct assessment may be a better choice. Take academic assessment as an example: If we wanted to assess how well a child reads, we could ask her to fill out a questionnaire in which she rates her own reading skills. But a sound direct assessment—in which she reads a text and answers questions about it, for example—is likely to be more informative. Similarly, we could ask a child to rate his own skill at reading facial expressions, but it may be better to directly assess the skill by showing him pictures of people with various facial expressions and asking him what the people are feeling.

In contrast to SEL thinking skills, behavioral skills are expressed outwardly, so they can be directly observed—for example, when a child compromises with or hits another child. Behavioral observation is designed to measure the frequency and intensity of socially positive and aversive behavior. However, infrequent but highly consequential events are less likely to be observed, and it's difficult to use observation in a way that yields reliable scores that are appropriate for their intended interpretation. Peer ratings can also assess socially positive and aversive behavior, and starting in late elementary school, questionnaires can ascertain children's own view of their social-emotional characteristics. But none of these methods is optimal for assessing social behavior. It's both time and labor intensive to get reliable and valid data from observation. Peer ratings are prohibitively complex to administer, score, and interpret. When completing self-report questionnaires, children may indicate a socially desirable response, whether or not it's accurate.

Two methods of assessing behavior are more feasible in schools than the rest. First, teacher rating scales can yield reliable and valid assessments of overall behavioral tendencies, and they're easy to use, score, and interpret. They have limitations, however. For example, different raters might judge the same child's behavior differently. Furthermore, rating scales place a burden on teachers, who may have to rate many students. A second approach, direct behavior ratings, retains the advantages of both direct observation (objectivity and behavior in naturalistic settings) and rating scales (simplicity and consistency). In this approach, a teacher rates the frequency of a small number of clear target behaviors (such as whether a child talks out of turn) over a brief period.[65] Direct behavior ratings have great potential for characterizing child behavior, screening for disruptive behavior problems, and monitoring progress.

Self-control includes specific thinking skills and their behavioral expression. The thinking dimensions of self-control may be measured through direct assessment and, in some cases, self-report. Behaviors that reflect the absence of self-control may be measured through observation, rating scales, or direct behavior ratings. Self-control may also be reflected in beliefs and attitudes about the self. When grit—a component of self-control defined as "perseverance

and passion for long-term goals"—has been measured through self-reporting, researchers have found that the scores are reliable and are associated with important outcomes.[66]

Summary

I've highlighted the prospect of identifying meaningful, measurable, and malleable SEL skills that correspond to state standards, as well as the existing tactics for assessing those skills. But no systems yet exist for large-scale assessment of SEL skills. Thus, SEL assessment is in its early stages. We have sufficient proof of concept to feel confident that we can create feasible, rigorous, and scalable assessment systems, but no systems developed so far meet schools' important and varied needs.

In addition to the SEL skills I've described, readers might see other dimensions of SEL as important. My list omits some often discussed constructs, such as growth mindsets.[67] Based on Carol Dweck's seminal work on children's implicit theories of intelligence, the concept of mindsets focuses on an important belief system—children's beliefs about the nature of intelligence. Mindsets are meaningful, measurable, and malleable, and are important and influential ideas with strong implications for the classroom. In our conception, however, mindsets (and other beliefs and attitudes) are distinct from the mental, behavioral, and self-control skills that make up what we call SEL.

I don't claim that the SEL skills I've reviewed in this article are the only ones that should be included in our shared understanding of SEL. But we must achieve sufficient consensus to guide what we measure; strongly consider matching the method to what is measured; use a developmental perspective to guide measurement; and make certain we're measuring the dimensions of SEL that are meaningful, measurable, and malleable. As we tackle the daunting task of creating assessments with the same rigor and sophistication as achievement tests, these principles will help us make great strides.

What a Serious SEL Assessment Effort Would Require

How much would it cost in money and other resources to develop SEL assessment systems that meet schools' educational needs? Though a precise estimate isn't feasible, our research group's work to develop and validate a web-based system to measure several SEL thinking skills may at least give an idea of the size of the investment that will likely be required.

My colleagues and I recognized that despite SEL's importance to learning, we had few tools to assess children's SEL thinking skills; most social-emotional assessments are designed to measure children's behavior. Yet rigorous assessment of SEL thinking skills is critical, not only because those skills are reflected in standards, but also because understanding children's social-emotional thinking skills can guide educators' instructional decisions. For example, if a child performed poorly on a social problem-solving test, teachers could use evidence-based instructional strategies to help her improve her social problem-solving skills.

Thus, we set out to create a web-based system—called SELweb—to assess SEL thinking skills in children from kindergarten to third grade. SELweb measures children's ability to recognize others' emotions, take others' perspectives, solve social problems,

and practice self-control. All of its modules are illustrated and narrated so that children as young as kindergarten age can complete the assessment independently. We also matched method to what we measure. The modules are direct assessments, where children complete challenging and engaging tasks that require them to demonstrate thinking skills. The case of SELweb illustrates what one SEL assessment system can do. But it also illustrates the tremendous effort required to create scalable, scientifically sound, usable, and feasible SEL systems.

To evaluate how well SELweb measured the target skills, we mounted two field trials with a large and diverse sample of 4,462 kindergarteners through third-graders. SELweb's score reliabilities, which index consistency of measurement, were comparable to well-developed achievement tests. In both field trials, scores on the assessment modules fit a hypothesized model of SEL thinking skills that includes four factors—emotion recognition, perspective-taking, problem-solving, and self-control. Overall, higher scores on SELweb were positively associated with teacher-reported social skill, peer acceptance, and academic competence, and negatively associated with teacher-reported problem behavior. In addition, scores on SELweb's different modules were more strongly associated with alternative measures of the same construct than they were associated with alternate measures of different constructs. These findings support the conclusion that SELweb scores reflect what they were designed to measure.[68] As a final step, we collected SELweb data from 4,419 students in six states to create age-based norms, so that a child's performance on SELweb can be judged in comparison to a national sample of children the same age.

If we want to get serious about assessing SEL, we'll need to invest significant resources.

All this took four years, considerable financial support from the Institute of Education Sciences, and many person-hours. SELweb demonstrates that it's possible to create engaging, scalable, scientifically sound, and useful SEL assessment systems. And yet like any assessment system, SELweb can't do all things. It measures thinking skills but not behavioral skills. Its design and psychometric properties make it appropriate for guiding classroom instruction and evaluating programs to foster SEL skills. It could perhaps be used for low-stakes monitoring of student progress toward some, but not all, SEL standards. Our experience tells us that if we want to get serious about assessing SEL, we'll need to invest significant resources and consider how to sustain and continually improve our assessments—much the same way that standardized achievement tests require large initial investments and continual upkeep.

What would it take, then, to create a developmentally appropriate, multimethod, multirater SEL assessment system for K–12? Consider the assessments developed to measure children's progress toward the Common Core educational standards.

In September 2014, *Education Week* reported that more than $300 million in contracts had been awarded to testing companies to develop assessment systems.[69] Those investments were directed to highly competent organizations with strong track records of rigorous academic assessment development. It seems likely that we would need a comparable commitment of resources to develop SEL assessment systems with the same rigor and utility.

Filling the Void

Promising assessments that measure SEL thinking skills, behavioral skills, and self-control skills exist or are in development. However, we have yet to invest enough resources to produce robust and scalable systems that correspond to state standards and that allow educators to use assessment to foster children's social-emotional development. Some existing and emerging tools are appropriate for formative assessment and program evaluation. However, they cover some dimensions of SEL better than others, and we have few options to achieve other assessment goals, such as monitoring children's progress toward meeting SEL standards.

How can we fill the gaps? First, SEL assessment development efforts should meet the highest ethical and scientific standards.[70] For most SEL assessment goals, that means going well beyond simple survey construction to developing multimethod, multirater systems that have been well constructed and rigorously evaluated. This will require a level of test development effort and rigor that has typically been reserved for achievement tests.

Second, developers should design and build SEL assessment systems specifically for educational use. Many existing tools were developed either for research (such as emotion-recognition tasks) or for clinical applications (such as most rating scales). Thus educators must retrofit these assessments for off-label uses. SEL assessments designed with educators in mind should be feasible to deploy in schools at scale; focus on strengths; use up as little instructional time as possible; and quickly and flexibly report informative results. As much as possible, such assessments should also measure dimensions of SEL that are reflected in state standards and in the best evidence-based SEL programs.

Third, developers should focus on measuring dimensions of SEL that span the three categories of thinking, behavioral, and self-control skills. They should also choose to measure skills that are meaningful, measurable, and malleable—that is, skills that are associated with important outcomes, can be assessed feasibly, and can be influenced by experience. In rare instances, however, we might want to measure a skill that's meaningful and measurable even though it isn't clear that the skill is malleable via instructional strategies. Measuring such skills could encourage researchers to develop curricular and instructional strategies to shape them.

Fourth, assessments methods should be matched to what's being measured. I believe that direct assessment is best for SEL thinking skills; rating scales and direct behavior ratings are best for behavioral skills; and a combination of direct assessment and rating scales is best for self-control skills. But those guidelines are debatable. What's most important is that developers thoughtfully pair their methods to what they're measuring. SEL assessments

covering all three categories of thinking, behavioral, and self-control skills will therefore need more than one method and more than one rater. Such a system would require direct assessment and teacher rating at a minimum, and might also include peer nominations and direct behavior ratings. Multimethod, multirater assessment systems will corroborate students' SEL skill levels, creating a hedge against outlier performance on any one measure.

Fifth, SEL assessments should be developmentally appropriate. SEL skills change with age, and research tells us what changes to expect. Broadly, SEL assessments should account for two kinds of developmental changes. The first involves constructs whose meaning and manifestation remain the same as children's performance improves with age. For example, although facial emotion recognition is the same skill throughout the lifespan, individuals become better at it as they grow older. To measure such skills, assessments should include items with a range of difficulty that corresponds to the variability in the skills of all children in the age range to be tested. The second kind of change involves constructs whose meaning remains the same but whose manifestation changes. For example, as children traverse middle childhood, in addition to recognizing emotions through their behavioral expression, they come to understand that people can have mixed emotions, such as being happy for a friend and sad for oneself, and moral emotions, such as guilt and pride.[71] This new kind of understanding is different from emotional understanding in very young children and therefore requires a different assessment method.

Last, the intended use of any SEL assessment system should be clearly specified from the design stage through the large-scale rollout—and before it's rolled out, the developers must be able to show sufficient evidence that the assessment is appropriate for that purpose. Any other uses should be clearly characterized as "off-label," and potential negative consequences of such uses should be described. The user's goals and practices can't be built into the assessment technology itself; rather, assessment developers must communicate appropriate use in documentation and training.

Excellent assessment is crucial to making progress on social-emotional learning, from policy to practice to research. How else can we know children's strengths and needs, and therefore, how to target instruction to foster character? How else can we know whether a set of practices works? How else can we know to what heights of character development students have risen? How else can we know whether our system of education has met state standards (assuming such standards apply to the education of character)? These are not idle questions. If nature abhors vacuums, educational fads feast on them. All of us—scientists, practitioners, parents, and policymakers—should hope that the best evidence of what works will lead to practices that nurture SEL skills. Assessment is the foundation for collecting such evidence.

ENDNOTES

1. Joseph A. Durlak et al. (eds.), *Handbook of Social and Emotional Learning: Research and Practice* (New York: Guilford Press, 2015).

2. Education Week Research Center, *Social and Emotional Learning: Perspectives from America's Schools* (Bethesda, MD: Editorial Projects in Education Inc., 2015).

3. Linda Dusenbury et al., *State Learning Standards to Advance Social and Emotional Learning: The State Scan of Social and Emotional Learning Standards, Preschool Through High School* (Chicago: CASEL, 2011).

4. "What is SEL?" Collaborative for Academic Social and Emotional Learning, accessed January 7, 2017, http://www.casel.org.what-is-sel/; "Core SEL Competencies," Collaborative for Academic Social and Emotional Learning, accessed January 7, 2017, http://www.casel.org/social-and-emotional-learning/core-competencies.

5. Jenny Nagaoka et al., *Foundations for Young Adult Success: A Developmental Framework* (Chicago: Consortium on Chicago School Research, University of Chicago, 2015).

6. National Research Council, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (Washington, DC: National Academies Press, 2012).

7. Stephanie M. Jones and Suzanne M. Bouffard, "Social and Emotional Learning in Schools: From Programs to Strategies," *SRCD Social Policy Report* 26 (2012): 3–33.

8. Elizabeth A. Lemerise and William F. Arsenio, "An Integrated Model of Emotion Processes and Cognition in Social Information Processing," *Child Development* 71 (2000): 107–18; Nagaoka et al., *Foundations*.

9. Meryl E. Lipton and Stephen Nowicki Jr., "The Social-Emotional Learning Framework (SELF): A Guide for Understanding Brain-Based Social-Emotional Learning Impairments," *Journal of Developmental Processes* 4 (2009): 99–115.

10. Frank M. Gresham, "Conceptual and Definitional Issues in the Assessment of Children's Social Skills: Implications for Classification and Training," *Journal of Clinical Child Psychology* 15 (1986): 3–15, doi: 10.1207/s15374424jccp1501_1.

11. Roy F. Baumeister, Kathleen D. Vohs, and Dianne M. Tice, "The Strengths Model of Self-Control," *Current Directions in Psychological Science* 16 (2007): 351–55; C. Cybele Raver, "Low-Income Children's Self-Regulation in the Classroom: Scientific Inquiry for Social Change," *American Psychologist* 67 (2012): 681–89, doi: 10.1037/a0030085.

12 Angela L. Duckworth, "The Significance of Self-Control," *Proceedings of the National Academy of Sciences of the United States of America* 108 (2011): 2639–40, doi: 10.1073/pnas.1019725108; 10.1073/pnas.1019725108; Radiah Smith-Donald et al., "Preliminary Construct and Concurrent Validity of the Preschool Self-Regulation Assessment (PSRA) for Field-Based Research," *Early Childhood Research Quarterly* 22 (2007): 173–87, doi: 10.1016/j.ecresq.2007.01.002.

13. Transforming Education, "How We Approach Partnerships," accessed March 9, 2017, https://www.transformingeducation.org/partnerships.

14. Carroll Izard et al., "Emotion Knowledge as a Predictor of Social Behavior and Academic Competence in Children at Risk," *Psychological Science* 12 (2001): 18–23.

15. Stephen Nowicki Jr. and Marshall P. Duke, "Individual Differences in the Nonverbal Communication of Affect: The Diagnostic Analysis of Nonverbal Accuracy Scale," *Journal of Nonverbal Behavior* 18 (1994): 9–35, doi: 10.1007/BF02169077.

16. Clancy Blair and Rachel P. Razza, "Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten," *Child Development* 78 (2007): 647–63, doi: 10.1111/j.1467-8624.2007.01019.x.

17. Sue Walker, "Gender Differences in the Relationship Between Young Children's Peer-Related Social Competence and Individual Differences in Theory of Mind," *Journal of Genetic Psychology* 166 (2005): 297–312, doi: 10.3200/GNTP.166.3.297-312; Robin Banerjee and Dawn Watling, "Children's Understanding of Faux Pas: Associations with Peer Relations," *Hellenic Journal of Psychology* 2 (2005): 27–45.

18. Susanne A. Denham et al., "Observing Preschoolers' Social-Emotional Behavior: Structure, Foundations, and Prediction of Early School Success," *Journal of Genetic Psychology: Research and Theory on Human Development* 173 (2012): 246–78, doi: 10.1080/00221325.2011.597457.

19. Clark McKown et al., "Social-Emotional Learning Skill, Self-Regulation, and Social Competence in Typically Developing and Clinic-Referred Children," *Journal of Clinical Child and Adolescent Psychology* 38 (2009): 858–71, doi: 10.1080/15374410903258934.

20. Ibid.

21. Illinois State Board of Education, "Social and Emotional Learning Standards," accessed January 7, 2017, https://www.isbe.net/Pages/Social-Emotional-Learning-Standards.aspx.

22. Nowicki and Duke, "Individual Differences."

23. Francisco Pons, Paul L. Harris, and Marc de Rosnay, "Emotion Comprehension Between 3 and 11 Years: Developmental Periods and Hierarchical Organization," *European Journal of Developmental Psychology* 1 (2004): 127–52, doi: 10.1080/17405620344000022

24. Henry M. Wellman and David Liu, "Scaling Theory-of-Mind Tasks," *Child Development* 75 (2004): 523–41, doi: 10.1111/j.1467-8624.2004.00691.x.

25. Robert L. Selman, *The Growth of Interpersonal Understanding: Developmental and Clinical Analysis* (New York: Academic Press, 1980).

26. Clark McKown et al., "Direct Assessment of Children's Social-Emotional Comprehension," *Psychological Assessment* 25 (2013): doi: 10.1037/a0033435.

27. Marit Korkman, Ursula Kirk, and Sally Kemp, *NEPSY®—Second Edition* (San Antonio, TX: Pearson Assessment, 2007); Janis B. Kupersmidt, Rebecca Stelter, and Kenneth A. Dodge, "Development and Validation of the Social Information Processing Application: A Web-Based Measure of Social Information Processing Patterns in Elementary School–Age Boys," *Psychological Assessment* 23 (2011): 834–47, doi: 10.1037/a0023621.

28. Clark McKown et al., "Web-Based Assessment of Children's Social-Emotional Comprehension," *Journal of Psychoeducational Assessment* 34 (2016): 322–38, doi: 10.1177/0734282915604564.

29. Joseph A. Durlak et al., "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions," *Child Development* 82 (2011): 405–32, doi: 10.1111/j.1467-8624.2010.01564.x; Justus J. Randolph and R. Shawn Edmondson, "Using the Binomial Effect Size Display (BESD) to Present the Magnitude of Effect Sizes to the Evaluation Audience," *Practical Assessment, Research, and Evaluation* 10 (2005): article 4.

30. Nicole M. Russo-Ponsaran et al., "Efficacy of a Facial Emotion Training Program for Children and Teenagers with Autism Spectrum Disorders," *Journal of Nonverbal Behavior* 40 (2016): 13–38, doi: 10.1007/s10919-015-0217-5; Candice Southall and Jonathan M. Campbell, "What Does Research Say About Perspective-Taking Interventions for Students with HFASD?" *Exceptional Children* 81 (2015): 194–208, doi: 10.1177/0014402914551740; Ashley M. Candelaria, Alicia L. Fedewa, and Soyeon Ahn, "The Effects of Anger Management on Children's Social and Emotional Outcomes: A Meta-Analysis," *School Psychology International* 33 (2012): 596–614, doi: 10.1177/0143034312454360.

31. James Clyde DiPerna, Robert J. Volpe, and Stephen N. Elliott, "A Model of Academic Enablers and Elementary Reading/Language Arts Achievement," *School Psychology Review* 31 (2002): 298–312; Frank M. Gresham, "Conceptual and Definitional Issues."

32. James Clyde DiPerna and Stephen N. Elliott, "Development and Validation of the Academic Competence Evaluation Scales," *Journal of Psychoeducational Assessment* 17 (1999): 207–25, doi: 10.1177/073428299901700302.

33. Kathryn R. Wentzel, "Does Being Good Make the Grade? Social Behavior and Academic Competence in Middle School," *Journal of Educational Psychology* 85 (1993): 357–64.

34. Gian V. Caprara et al., "Prosocial Foundations of Children's Academic Achievement," *Psychological Science* 11 (2000): 302–6, doi:10.1111/1467-9280.00260; Damon E. Jones, Mark Greenberg, and Max Crowley, "Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness," *American Journal of Public Health* 11 (2015): 2283–90, doi: 10.2105/AJPH.2015.302630.

35. Deborah Stipek and Sarah Miles, "Effects of Aggression on Achievement: Does Conflict with the Teacher Make it Worse?" *Child Development* 79 (2008): 1721–35, doi: 10.1111/j.1467-8624.2008.01221.x.

36. Stephen P. Hinshaw, "Externalizing Behavior Problems and Academic Underachievement in Childhood and Adolescence: Causal Relationships and Underlying Mechanisms," *Psychological Bulletin* 111 (1992): 127–55, doi: 10.1037/0033-2909.111.1.127.

37. Kenneth A. Dodge, "Behavioral Antecedents of Peer Social Status," *Child Development* 54 (1983): 1386–99.

38. Jones et al., "Early Social-Emotional Functioning."

39. Illinois State Board of Education, "Social and Emotional Learning Standards."

40. Frank M. Gresham and Stephen N. Elliott, *Social Skills Improvement System: Rating Scales* (Bloomington, MN: Pearson Assessments, 2008).

41. Paul A. LeBuffe, Valerie B. Shapiro, and Jack A. Naglieri, *Devereux Student Strengths Assessment, K–8th Grade* (Renton, WA: Apperson, 2012).

42. James C. DiPerna and Stephen N. Elliott, *Academic Competence Evaluation Scales* (San Antonio, TX: The Psychological Corporation, 2000).

43. Durlak et al., "Impact"; Randolph and Edmondson, "Binomial Effect Size Display."

44. Joseph A. Durlak and Anne M. Wells, "Evaluation of Indicated Preventive Intervention (Secondary Prevention) Mental Health Programs for Children and Adolescents," *American Journal of Community Psychology* 26 (1999): 775–802, doi: 10.1023/A:1022162015815.

45. Catherine P. Bradshaw, Tracy E. Waasdorp, and Philip J. Leaf, "Effects of School-Wide Positive Behavioral Interventions and Supports on Child Behavior Problems," *Pediatrics* 130 (2012): 1136–45, doi: 10.1542/peds.2012-0243.

46. Catherine P. Bradshaw, Tracy E. Waasdorp, Philip J. Leaf, "Examining Variation in the Impact of School-Wide Positive Behavioral Interventions and Supports: Findings from a Randomized Controlled Effectiveness Trial," *Journal of Educational Psychology* 107 (2015): 546–57, doi: 10.1037/a0037630.

47. Daniel H. Tingstrom, Heather E. Sterling-Turner, and Susan M. Wilczynski, "The Good Behavior Game: 1969–2002," *Behavior Modification* 30 (2006): 225–53.

48. Clancy Blair and C. Cybele Raver, "School Readiness and Self-Regulation: A Developmental Psychobiological Approach," *Annual Review of Psychology* 66 (2015): 711–31, doi: 10.1146/annurev-psych-010814-015221.

49. Roopa V. Iyer et al., "Peer Victimization and Effortful Control: Relations to School Engagement and Academic Achievement," *Merrill-Palmer Quarterly: Journal of Developmental Psychology* 56 (2010): 361–87, doi: 10.1353/mpq.0.0058.

50. Angela L. Duckworth and Martin E.P. Seligman, "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents," *Psychological Science* 16 (2005): 939–44, doi: 10.1111/j.1467-9280.2005.01641.x.

51. Terri E. Moffitt et al., "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety," *PNAS: Proceedings of the National Academy of Sciences of the United States of America* 108 (2011): 2693–98, doi: 10.1073/pnas.1010076108.

52. Illinois State Board of Education, "Social and Emotional Learning Standards."

53. Angela L. Duckworth, Tamar S. Gendler, and James J. Gross, "Self-Control in School-Age Children," *Educational Psychologist* 49 (2014): 199–217.

54. Yuichi Shoda, Walter Mischel, and Philip K. Peake, "Predicting Adolescent Cognitive and Self-Regulatory Competencies from Preschool Delay of Gratification: Identifying Diagnostic Conditions," *Developmental Psychology* 26 (1990): 978–86, doi: 10.1037/0012-1649.26.6.978.

55. Smith-Donald et al., "Preliminary Construct and Concurrent Validity."

56. Susanne A. Denham et al., "Factor Structure of Self-Regulation in Preschoolers: Testing Models of a Field-Based Assessment for Predicting Early School Readiness," *Journal of Experimental Child Psychology* 111 (2012): 386–404, doi: 10.1016/j.jecp.2011.10.002.

57. This task was drawn from the work of Jonna Kuntsi et al., "Test–Retest Reliability of a New Delay Aversion Task and Executive Function Measures," *British Journal of Developmental Psychology* 19 (2001): 339–48, doi: 10.1348/026151001166137.

58. This task was drawn from the work of Paraskevi Bitsakou et al., "Probing the Limits of Delay Intolerance: Preliminary Young Adult Data from the Delay Frustration Task." J*ournal of Neuroscience Methods* 151 (2006): 38–44, doi: 10.1016/j.jneumeth.2005.06.031.

59. Andy Wright and Adele Diamond, "An Effect of Inhibitory Load in Children While Keeping Working Memory Load Constant," *Frontiers in Psychology* 5 (2014): 1–9, doi: 10.3389/fpsyg.2014.00213.

60. Jessica D. Burdick, Amanda L. Roy, and C. Cybele Raver, "Evaluating the Iowa Gambling Task as a Direct Assessment of Impulsivity with Low-Income Children," *Personality and Individual Differences* 55 (2013): 771–76, doi: 10.1016/j.paid.2013.06.009.

61. Duckworth and Seligman, "Self-Discipline Outdoes IQ."

62. C. Cybele Raver et al., "CSRP's Impact on Low-Income Preschoolers' Preacademic Skills: Self-Regulation as a Mediating Mechanism," *Child Development* 82 (2011): 362–78, doi: 10.1111/j.1467-8624.2010.01561.x.

63. Kimberly A. Schonert-Reichl et al., "Enhancing Cognitive and Social-Emotional Development through a Simple-to-Administer Mindfulness-Based School Program for Elementary Students: A Randomized Controlled Trial," *Developmental Psychology* 51 (2015): 52–66, doi: 10.1037/a0038454.

64. McKown et al., "Web-Based Direct Assessment."

65. Sandra M. Chafouleas, "Direct Behavior Rating: A Review of the Issues and Research in Its Development," *Education & Treatment of Children* 34 (2011): 575–91, doi: 10.1353/etc.2011.0034.

66. Angela L. Duckworth and Margaret L. Kern, "A Meta-Analysis of the Convergent Validity of Self-Control Measures," *Journal of Research on Personality* 45 (2011): 259–68, doi: 10.1016/j.jrp.2011.02.004.

67. Carol Dweck, *Mindset: The New Psychology of Success* (New York: Random House, 2006).

68. Clark McKown, "Challenges and Opportunities in the Direct Assessment of Children's Social-Emotional Comprehension," in *Handbook of Social and Emotional Learning*, ed. Joseph A. Durlak et al. (New York: Guilford Press, 2015), 320–35.

69. Sean Cavanagh, "Common-Core testing Contracts Favor Big Vendors," *Education Week*, September 30, 2014, http://www.edweek.org/ew/articles/2014/10/01/06contract.h34.html.

70. For example, those outlined in American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, *Standards for Educational Testing* (Washington, DC: American Educational Research Association, 2004).

71. Pons, Harris, and de Rosnay, "Emotion Comprehension."