Key Design Principles for Direct Assessments of SEL:
Lessons Learned from the 2017-2019 Design Challenges

Background

Measuring students' social and emotional competencies has become a priority for school districts and out-of-school-time programs that are committed to fostering social and emotional learning (SEL)[1]. While surveys of students' self-perceptions and teachers' observations are common and widely available, direct assessments, in which students directly demonstrate their level of competency, are largely missing from the catalog of assessment tools appropriate for educational settings. Direct assessments are well-suited to measuring dimensions of social and emotional learning that are not easily observed unless the individual has a direct opportunity to perform them, such as the thinking skills involved to understand what other people are thinking and feeling (i.e., social perspective taking). Some of these skills and behaviors are not easily inferred from a person's general actions, making them a challenge for observers to assess. As a result, more work is needed to create and test direct assessments for application in educational settings[2].

AUTHORS

Lindsay Read
University of Chicago
(formerly CASEL)

Noah Bookman
CORE Districts

Whitney Stormberg
CORE Districts

[1] Social and emotional learning (SEL) has been defined in various ways. One of the most commonly cited models, that of the Collaborative for Academic, Social, and Emotional Learning (CASEL), defines SEL as "the process through which students 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." That process incorporates competencies that include knowledge, skills, and attitudes that affect the quality of students' interactions with one another and their ability to adapt to and successfully navigate the complexities of daily life.

[2] McKown, C., Read, L., & Bookman, N. (2017). *Key Design Principles for Direct Assessments of SEL: Lessons Learned from the First Design Challenge.* https://measuringsel.casel.org/wp-content/uploads/2017/09/AWG-Design-Challenge-Direct-Assessments-of-SEL.pdf

Building on lessons learned from three annual design challenges, this brief aims to stimulate the development and continuous improvement of direct assessments of student SEL by laying out a set of design principles. These principles are based upon an examination of available and proposed direct assessment tools that were nominated and reviewed during the three-year Design Challenge. The primary audiences are assessment developers and practitioners in districts, schools, and out-of-school-time settings who might use these design principles as a framework for evaluating and selecting SEL direct assessments for incorporation into their assessment systems. Ultimately, we hope this brief can contribute to a conversation across the educational assessment community that focuses upon the creation and effective use of SEL direct assessments within a broader assessment system that supports instruction and positive student development.

What do we mean by direct assessments?

Direct assessment refers to a method of assessment in which students demonstrate their competence by solving challenging social and emotional tasks. Direct assessments differ from most of the current generation of SEL competence assessments in which students, educators, and parents respond to surveys or make ratings based on general observations. Most existing SEL assessments rely on a reporter who judges behaviors or tendencies. Teacher reports, for example, typically measure a teacher's view of the frequency of student behaviors they see in general and not based on specific opportunities to observe the demonstration or performance of the competence directly. This method is well-suited to assessing the frequency of clearly observable behaviors. Self-report is well-suited to assessing attitudes in students old enough to understand questionnaires. However, both teacher report and self-report are susceptible to rater biases. Teachers may have different reference groups in mind when rating a student, and students may respond in ways that reflect social expectations more than actual skill level when completing self-report questionnaires (McKown, Read, & Bookman, 2017).

The Design Challenge Process

The design principles arose from studying the 18 winners of three annual Design Challenges focused on direct assessments, the first in 2017, the second in 2018, and the third in 2019. The approach to and composition of the first set of assessments informed the approach to the second, and the second changed the approach used in the third.

In 2017, the Assessment Work Group's Design Challenge team held a public competition to scan the landscape of SEL direct assessments and uncover promising ideas. As such, our search was broad and included a solicitation for any direct assessments of SEL, even if they were only in the conceptual stage of development. Specifically, the committee sought submissions that were: (1) direct assessments of SEL, including all forms of assessment in which social-emotional competence is measured from a child's performance on a challenging task, (2) innovative, and (3) aimed at any age range between pre-kindergarten and 12th grade. A review panel of assessment experts evaluated the 20 submissions we received according to a set of criteria we developed based upon the existing literature on direct assessments[3]. Many excellent ideas and concepts emerged from the

---

[3] Submissions were evaluated based upon the following 8 criteria: (1) the innovativeness of the assessment, (2) the clarity of the assessment goals, (3) the usefulness of the assessment data for teaching and learning, (4) the potential of the assessment to be administered at scale, (5) the usefulness of the data reporting, (6) the developmental and cultural appropriateness of the assessment, (7) the potential of the assessment to be engaging to users, (8) the assessment's technical properties, or, for early stage concepts, the potential for those technical properties to be evaluated.

competition, yet most were not yet useful to practitioners within educational settings. This is not a commentary on the quality of the submissions or ideas, but rather on the state of the field of direct assessment development and the challenge of designing tools to measure SEL skills directly, especially skills that are not easily observable without more explicit opportunities to demonstrate specific competencies. Descriptions of the 2017 winners are available in Appendix A and in more detail in our 2017 brief.

To address the desire to find assessments that could be implemented in educational settings, in the 2018 cycle we invited practitioners to share their assessment needs. From the over 60 submissions we received from practitioners, 10 were selected based upon these criteria: their submission applied to PreK-12th grade students and provided a clear and actionable description of need that could be addressed through a direct assessment of social-emotional competencies. We then asked assessment developers to submit proposals for how direct assessments could address these specific practitioner-generated needs. We asked that submitters identify one of the needs to respond to and describe their assessment and how it would address the stated need. We emphasized that submissions could be designs, prototypes, or fully developed assessments. In response, we received 11 submissions—six from practitioners, two from researchers, two from test developers, and one from a consultant. Six submissions were in the design phase, four reflected working prototypes, and one was a fully developed assessment. Judges selected five submissions as winners. The winning submissions included game-based, computer simulations to assess elementary-aged students' social information processing, and a text-based game to assess the role school values play in high school students' educational pathways. While several of these submissions were appropriate for school use, most had not been fully implemented. Descriptions of the winners are available in Appendix B and in more detail in our 2018 brief.

For the final cycle in 2019, we prioritized those assessments that had demonstrated success in educational settings. As we learned through our 2017 and 2018 challenges, assessments need to be useful to practitioners. One way to ensure this is to study those that have been successfully used in practice. Our 2019 design challenge broadened its scope beyond direct assessment to include examples of assessments that were implemented in ways that could be models for others and could serve to inform the design principles. Therefore, instead of strictly focusing on direct assessments, we opened our search to well-defined observational and other direct rubrics in use in practice. These approach direct assessment in that they include clearly defined, often highly detailed, learning progressions against which students' can demonstrate their competence.

Through this broader scope, we identified six assessments already in use in schools and out-of-school-time programs primarily through the Assessment Work Group members, the larger international Collaborator Network, and related communities. To learn more about the assessments and narrow down qualifying exemplars, we contacted developers and asked them what the assessment measures, how many schools use the assessment, what the SEL assessment data shows, and how the findings from the data are used in practice. From the 10 assessments that were submitted and reviewed, six exemplars were selected based on the strength of their SEL measures, the directness of student engagement in the assessment, assessment implementation of roughly a year, use of SEL data for direct classroom/student group improvement, and non-winners of previous design challenges. Two of the six exemplars are web-based direct assessment games measuring performance on virtual tasks, while the other four use observational rubrics based upon various learning progressions. We interviewed each of the developers and at least one user of each assessment at the school or district level to learn how they use the data from the assessments. The exemplars are described briefly in the next section and more fully in Appendix D.

The 2019 Exemplars

The 2019 Design Challenge identified six measures based upon their successful implementation in educational settings. They fall into three categories based upon ways in which they vary from one another: two are web-based assessments that use virtual scenarios to assess performance, two are rubric-based assessments created within educational settings, and two are rubric-based assessments created within educational programs. Below are short descriptions of each, by category. More detailed descriptions are available in Appendix D.

Web-based assessments that use virtual scenarios to assess performance

The two measures in this first category provide opportunities for youth to directly demonstrate their social-emotional competencies/skills in a virtual situation presented online. These scenarios are focused on providing an opportunity for youth to directly demonstrate use of a particular set of competencies that the tool is designed to measure as they play the game. In these two exemplars, the opportunities to demonstrate a skill are built into the virtual scenarios and not dependent on the program or curriculum being used to teach SEL.

The **Minnesota Executive Function Scale (MEFS App)**, developed by Reflection Sciences, is a direct assessment designed to measure the following executive function skills in children as young as 24 months: working memory, inhibitory control, and cognitive flexibility. The developers' aim was to develop an intuitive assessment that could provide accurate and necessary data identifying developmental delays in neurocognitive skills, known as executive function skills, to help individuals manage emotions, demonstrate empathy, and achieve academic success. While many in the field believe executive function skills are distinct from social-emotional skills, we include the MEFS App because it is one of few available examples of a direct measure in use in educational settings. The MEFS App is "played" using a game-based application, often on a tablet, and takes about five minutes for a student to complete with the aid of a classroom teaching assistant. To play the game, students must place animal cards into one of two boxes on the screen. For example, in level 5 of game play, a child may be asked to place all cards with green lions in the green lion box and all cards with orange cheetahs in the orange cheetah box for a couple rounds until the instructions change. As students advance to the next level of difficulty, the rounds for one type of instruction are shorter, causing the child to switch rapidly back and forth between various games and instructions within a level. This gradual switch and add-on instructional technique for each level of game play makes the assessment sensitive enough to identify where a child has difficulty following the rules of the game, indicating more precisely where they are developmentally. The assessment is used in more than 200 locations, including 38 U.S. states and 18 countries, and is available in 14 languages. At one school in a Minnesota school district, kindergarten teachers use the data to learn about incoming preschoolers' executive function skills as they enter kindergarten. Students and families are informed that the assessment provides a fuller picture of the student's skills and where they are in their learning, similar to state assessments in other subjects. The MEFS App cloud-based platform is capable of reporting data automatically at the individual, group, and school level.

SELweb, developed by xSEL Labs, is a direct assessment designed to measure emotion recognition, self control, social perspective-taking, and social problem solving in students in kindergarten through 6th grade[4]. SELweb is "played" using a web-based platform. Students engage in several performance task segments including a facial recognition task (K3 only) that asks students to identify how a person feels based on their facial expression and a scenario on taking another person's perspective to identify how they are feeling in that moment. Students in grades K-3 can complete SELweb EE on their own in about 25 minutes. Students in grades 4-6 can complete SELweb LE on their own in about 30 minutes. SELweb provides score reports at the individual, classroom, grade (within school), school/program, and district levels and includes scores for the individual domains as well as an overall SEL competence score. More than 200 schools across the country have used SELweb to measure SEL outcomes and identify trends. In one Illinois elementary school, using the fall baseline data, teachers and staff are able to see the strengths and needs of their students and classrooms. Teacher groups meet with administrators and instruction coaches to discuss gaps and areas of concern. Coaches offer alignment guides with the Second Step SEL program. Second-grade teachers also use the data to pair students who aren't proficient in SEL with those that are, as part of a buddy program.

Rubric-based assessments created within school districts

Both of these assessments were created within districts with collective input from numerous stakeholders. As such, they are examples of assessments customized to a particular population and set of SEL values and competencies. The assumption is that because the district is emphasizing these skills, it is also providing ample opportunities for youth to demonstrate their skills and for trained observers to assess different levels of competency. In contrast to the virtual scenario-based assessments in the first category, the districts are responsible for creating the specific situations in which students are able to demonstrate their skills. How well this is done depends, of course, on implementation, which is not a focus of this brief.

These rubric-based assessments will be most useful when there is strong alignment between (1) the language used to define and talk about the skills and (2) the opportunities students have to demonstrate the skills in the course of their work with those who are rating their competence (e.g., teachers). It is this strong alignment that makes these measures more direct than common sets of teacher ratings about frequencies of general behaviors.

The 5 Scholarly Habits, developed by Two Rivers Public Charter School District, is an observational rubric that teachers fill out to measure students' SEL competencies according to the following statements: I know myself; I am independent and resilient; I show compassion and embrace diversity; I can connect and collaborate; and I act with integrity. Students ages 3 to 14 receive developmentally appropriate classroom instruction on the five habits and are graded on them twice a year in addition to completing self-evaluations and short reflections on their performance in one skill area at the end of the week. Using common language across all grade levels and classrooms, these habits are embedded in the districts' culture from classroom instruction to small student group mentoring, recognition "swag," and school assemblies. The observational rubric contains standards for each scholarly habit. Each standard has corresponding "look fors" on a 1-5-point scale that helps teachers identify to students what they are regularly looking for in class. The standards and "look fors" were developed by teachers who worked together in grade-level bands to ensure the habits are consistent and recognizable to children as they receive instruction on them throughout the year and as they grow older and change grades. The SEL data generated from using the rubric helps inform decisions around how habits are taught and demonstrated in the classroom. The data also contributes to identifying trends, setting goals, making curriculum content decisions, and identifying student supports. Two Rivers' two elementary schools (PreK, age 3 to 5th grade) and one middle school (6th to 8th grade) serve a total of 850 students in the Washington, D.C. area.

---

[4] The developer of SELweb, Clark McKown, co-led the 2017 and 2018 Design Challenge but was not involved in the review or selection process in 2019.

The Standards Based Report Card, developed by San Francisco Unified School District (SFUSD), displays the scores teachers give students in academic subjects and SEL. Students receive the Standards Based Report Card three times a year. SFUSD teachers use a four-point observational rubric to measure social awareness, self-management, growth mindset, and self-efficacy of students in Transitional Kindergarten through 5th grade. The decision to focus upon these four competencies emerged from the four SEL measures they use through their membership in the CORE Districts. To support this assessment, SFUSD created a teacher guide to aid their teachers in creating safe and supportive classroom environments through the four SEL skills. Approximately 30 schools in SFUSD are grant-funded (10 new schools each year) to explicitly teach SEL in the classroom and assess students' SEL development on the four-point scale. Teachers use school site data from the previous year to help determine classroom next steps. They also use the SEL rubric to communicate one-on-one with students and families at parent-teacher conferences and provide parents and guardians translated family guides containing tools to support SEL development at home. Schools use the SEL data alongside culture and climate data for goal setting and planning.

Rubric-based assessments created within programs

These assessments were created within programs designed to build SEL competencies and the assessment functions as one part of a larger set of tools and activities. Given the tight alignment between program and assessment in this category, adopting the assessments alone without the program would likely not be useful. Below we describe the assessments within the context of their programs and provide an example of how a school or district is using it.

The Dream Life Skills Assessment Scale, developed by Dream a Dream, is an observational assessment of 4th-9th-graders that assesses the following five life skills: interacting with each other, solving problems and overcoming difficulties, taking initiative, managing conflict, and understanding instructions. Dream a Dream's mission is to offer "life skills as a solution to the challenge of failure to thrive stemming from growing up in the context of adversity." [5] The assessment is a part of Dream a Dream's art- and sports-focused After School Life Skills Programme, initially offered in partnership with schools in India. Program facilitators conduct pre and post assessments with each student at the beginning and end of the program/academic year in addition to daily session reports. Facilitators then use the SEL data to guide their After School Life Skills Programme activities. Once baseline data is submitted by facilitators and analyzed by the organization, facilitators receive individual child data, batch data, and schoolwide data that is also compared to developmental norms so facilitators know where each child is developmentally and can design specific interventions for them throughout the year. At the end of the academic year, facilitators do end-of-year observational assessments for each student, rating their improvement on a scale of 1 to 5. These scores along with two to three student stories of change for each grade (total of 12-18 stories per school) are submitted to Dream a Dream's Monitoring and Evaluation team. The organization uses the assessment data, along with individual student stories of change, to assess the program's impact, make improvements, update funders, and further research.

The Dream Life Skills Assessment Scale originated in India and is an open source tool used in 25 schools and two community centers in Bangalore and at least six nonprofits across six Indian cities. Globally, the assessment is known to be used in at least four nonprofits in four countries. Dream a Dream's Life Skills work also extends to its Career Connect workshops that prepare older students (ages 15-20) for the workforce and to its Teacher Development programs.

---

[5]   4th Wheel Social Impact, *Tracing the Impact of a Life Skills Programme on Quality of Life of Indian Adolescents*, Research Report. March 16, 2017.

The Deep Learning Progressions, developed by New Pedagogies for Deeper Learning (NPDL), is used by teachers to measure each student's developmental progression in the six Deep Learning Competencies, or 6Cs: communication, critical thinking, collaboration, creativity, citizenship, and character. There are six Deep Learning Progressions, one for each competency, wherein each progression there are highlights of various dimensions of the competency, each of which has a five-point rating criterion describing what a student needs to exhibit in order to receive that grade point. The five-point scale ranges from limited evidence to emerging, developing, accelerating, and proficient. Teachers use the Deep Learning Progressions as direct measures of students' development of each competency. To aid their grading, assessment evidence such as assignments, projects, tests, conversations, and observations are used to map student ratings along the progressions. In some schools, teachers have students use the Deep Learning Progressions to identify and track 6C growth of literary characters and even self-assess themselves to aid in student recognition and education of the 6Cs.

Eight countries, five U.S. states, and over 1,300 schools have used the 6Cs Deep Learning Progressions to foster inquiry-based learning. Districts within the U.S. have used the 6Cs data for schoolwide planning and to support discussions around best practices and classroom improvements. For example, one district uses the 6Cs rubrics to support deep learning in grades K-8. Schools use the previous year's assessment data to plan for the coming year. Both the district and its schools decide how many, and which, of the 6Cs to focus on as a community. Teachers use their student assessment data along with their deep learning design rubric and self-assessments to discuss best methods of practice and brainstorm deep learning improvements to make in the classroom.

We applaud all of the exemplar submissions and especially those 18 selected as exemplars throughout the three years of the Design Challenge. Together, they contributed significantly to the generation of the design principles described below. Each of the design challenge briefs published in 2017, 2018, and 2019 describes groundbreaking work in SEL performance assessment.

We surveyed developers of each of the 18 exemplars to find out how widely they are in use and we found that:

- Among the 2017 exemplars, at least one has been fully developed and has been used by thousands of students across multiple countries.

- Among the 2018 exemplars, at least two are fully developed and used by thousands of students across multiple countries.

- Among the 2019 exemplars, all six have all been fully developed and used* by thousands of students in schools, districts, and programs around the world. However, only one is a true direct measure of SEL competence. *Note that this was a condition of their selection in 2019.

The ultimate purpose of the Design Challenge was to support the generation and illustration of design principles and stimulate further development and adoption of these types of assessments. The principles that follow draw from lessons learned during the three years of the Design Challenge. The examples we use to illustrate them emerged through our strong emphasis on practical use in 2019. Given the relatively limited number of direct assessments of SEL competence in use in practice, the field could benefit from using the design principles to provide direction in assessment-related continuous improvement efforts before scaling in significant ways[6].

---

[6] Readers looking for more information on SEL assessment may be interested in a report released in October 2019 by the Assessment Work Group, the group responsible for this Design Challenge. This report provides more detail on the current state of the field of SEL assessment and offers a vision for the future.

The Design Principles

This section presents the design principles that emerged from the three-year design challenge and the interim design principles from years 1 and 2 (see interim principles in Appendix E). We also referenced the Guiding Principles for Equitable Performance Assessment Systems published by the California Performance Assessment Collaborative (see Appendix F). The principles emerged from a study of existing direct assessments and an evaluation by a panel of expert judges with input from practitioners in the field (see section above on the design challenge process). We acknowledge that there are a wide range of uses of direct assessments in educational settings and that some of these principles may be more valuable in certain contexts than in others. Nevertheless, strong direct assessments will be useful when they generally reflect most or all of these principles. Most of the principles include an example from our 2019 study of exemplars to illustrate how the principle may show up in practice. Our aim was to provide examples that were the best fit for each principle. As such, not all of the 2019 exemplars are represented, and in some cases where relevant information was not available, no example is included.

Design Principles for direct assessment of SEL	
Eight Design Principles for Direct Assessments:	
1	Direct Measure of Competence
2	Educationally Appropriate
3	Developmentally Appropriate
4	Culturally Appropriate
5	Technically Appropriate
6	Practical to Implement
7	Data are Useful
8	A Good Fit Overall

PRINCIPLE #1: Assessment is a Direct Measure of Competence

The assessment should directly measure students' demonstrated levels of competence in each targeted skill area.

In direct assessments, students demonstrate their social and emotional competencies through solving challenging social and emotional tasks. These assessments fall on a continuum of more or less direct measures of students' competence in a domain of SEL. How directly students demonstrate the competence and whether they do it in real or virtual situations can vary, but the intent is for the measure to capture the extent to which the student directly demonstrates a competence.

To be a direct measure, the assessment should meet these two criteria. First, the assessment materials (e.g., instructional/training materials and data reports) should define what the competence in each skill area looks like when demonstrated and, ideally, what different levels of competence look like along a developmental progression. Second, the measure or the context in which it is used should provide direct opportunities to demonstrate the competencies. These could be virtual opportunities, curriculum-based opportunities, or opportunities that are based in the program itself. For example, SELweb's platform provides virtual opportunities for students to demonstrate their competencies as they go through and react to the simulations and activities.

PRINCIPLE #2: ASSESSMENT IS EDUCATIONALLY APPROPRIATE

The assessment measures competencies that are intentionally being developed and taught, and the resources needed to assess them are appropriate for the educational setting.

Educationally appropriate assessments measure the learning that is occurring to build desirable SEL knowledge, skills, and behaviors in school and out-of-school-time settings. They should, in turn, be able to report these data in ways that can be used in these settings. To be educationally appropriate, assessments should be able to measure the SEL skills that educators want to build in students. Below we describe two ways this can occur: (1) by adopting an assessment to measure the competencies that a program is designed to teach and (2) by adopting a program that includes assessment. Note: There may be more ways this can occur than those we describe below.

- **Adopting an assessment to measure the competencies that a program is designed to teach:** There are a variety of program-agnostic assessments that measure the competencies programs are designed to teach. Combining these assessments and programs can be very powerful because the assessment data can guide instructional decision-making and measure student progress. SELweb, for example, is used in dozens of districts, many of which use an evidence-based SEL program such as Second Step. SFUSD also uses Second Step. While neither SELweb nor the SFUSD Standards-Based Report Card was developed specifically for Second Step, they are educationally appropriate assessments as long as they can measure the competencies the program aims to teach.

- **Adopting a program that includes assessment:** Some assessments demonstrate educational appropriateness by being directly aligned with an SEL curriculum or program. Very few programs build in assessment, but one example is the Dream Life Skills Assessment Scale, which demonstrates educational appropriateness because the assessment tools (e.g. student self-reflection, group-time discussion, stories of change) are embedded in a program that aims to build the very skills the assessment measures. It is possible, therefore, to measure competencies more directly when the program provides opportunities for students to demonstrate them and uses an assessment that measures those competencies.

PRINCIPLE #3: ASSESSMENT IS DEVELOPMENTALLY APPROPRIATE

The assessment has a clear developmental progression and is appropriate for the population for which it is used.

Developmentally appropriate assessments take into account the complex and nonlinear developmental trajectory along which a person learns and demonstrates SEL skills, knowledge, and behaviors. Putting in place age-appropriate performance tasks or learning progressions for grade levels or grade bands can ensure that students are assessed within a range of skill levels that is appropriate for where they are developmentally. Also, reports that display results within age-appropriate benchmarks can support the creation of developmentally appropriate school and classroom goals. To be developmentally appropriate, assessments should provide results that are relatively precise in their indication of a student's level of development along a range of skill level. For example:

- Two Rivers Public Charter School District's 5 Scholarly Habits assessment uses grade- level bands based upon similarities in students' developmental stages, e.g., cognitive abilities and classroom routines. Grade-level bands are as follows: Preschool and PreK (ages 3 and 4); kindergarten to 2nd grade; 3rd to 5th grade; and 6th to 8th grade. Teachers and staff intentionally worked together across grade level bands to create common SEL language and teaching dialogues that would be valued and understood by all students. To teach "I show compassion and embrace diversity," teachers observed what was going on in their classrooms and identified situations they could use as teachable moments. While the SEL language stays the same, the assessment's observational cues change by grade level band so teachers know which specific behaviors their students should be demonstrating at their particular developmental stage.

PRINCIPLE #4: ASSESSMENT IS CULTURALLY APPROPRIATE

The assessment is designed for the populations for which it is used.

Culturally appropriate assessments have been designed for—or are able to be adapted to—the cultural profiles of students (e.g., race/ethnicity, gender, socioeconomic background). They measure competencies that are valued in the cultures represented and are sensitive to differences in values across cultural groups. Along a spectrum of cultural appropriateness, an assessment would be more or less culturally appropriate depending on the degree to which it was designed for—or are able to be adapted to—the cultural profile of the population for which it is used.

We have limited evidence from our exemplars about the extent to which they have been designed for, or adapted to, specific populations. One example that touches on cultural appropriateness is Dream a Dream's After School Life Skills Programme, which was designed for children from vulnerable backgrounds in India who had little to no access to extra curriculars. Dream a Dream made specific additions to their program and pre-assessment based upon the needs of this population. The pre-assessment solicits information on how old each child looks, in addition to their actual age. Dream a Dream captures these data because they have found that students who do not look their actual age (either younger or older), tend to require more emotional support from the facilitator than other students.

We direct readers to the very thoughtful work on communicating the importance and ways of doing culturally responsive SEL assessment in the MeasuringSEL brief *Toward Transformative Social and Emotional Learning: Using an Equity Lens* and its summary brief, *Social and Emotional Learning: A Cultural Analysis.* The authors of these briefs point out that existing SEL assessments do not fully capture the complexity of what cultural identity means to individual students. For example, "a student who has experienced any kind of discrimination may use coping skills or code-switching to navigate social relationships. Some existing validated assessments, especially in the social psychology and identity literature, include perception items that capture students' cultural assets in a sensitive way."[7] As stated in the summary:

"[P]ractitioners and researchers must conceptualize, implement, and assess SEL in a way that adapts to students' cultural assets and recognizes their inherent strengths. In doing so, students may feel more respected and valued for who they are as individuals and experience a sense of belonging that is foundational to their learning success. Most existing SEL frameworks and programs [and assessments] have not yet made explicit the connections between students' cultural assets and their social and emotional development. However, as research on the intersection of SEL and equity has emerged, so have conversations for revising [these].

To ensure that practitioners and researchers measure SEL appropriately and ethically, assessment tools must be designed in a culturally responsive, strengths-based way. As the field of practical SEL measurement advances, leaders in the field should make considerations to capture cultural assets, consider contextual factors, and focus on adult SEL." *Note: These considerations are described in the underline detailed brief and summary brief.*

---

[7] Saavedra, E. & Nolan, E. (2018). Social and Emotional Learning: A Cultural Analysis. Adapted from Toward Transformative Social and Emotional Learning: Using an Equity Lens, 2018, by Jagers, R., Rivas-Drake, D., and Borowski, T.

PRINCIPLE #5: ASSESSMENT IS TECHNICALLY APPROPRIATE

Assessment is a valid and reliable measure of its specific uses.

Some of the strengths of direct assessments include their potential to measure social and emotional competence objectively and to yield highly reliable scores that are valid for a variety of purposes. They are also less vulnerable to rater biases than self-report or teacher rating scales. At the same time, they are comparatively expensive and complex to create and validate.[8] Yet it is still extremely important to adopt technically appropriate assessments and to understand what this means.

Reliability refers to the consistency of measurement. Validity refers to evidence that an assessment is appropriate for a particular use, or to make specific decisions. Validity is not a property of the instrument itself. Rather, it is a body of evidence supporting specific ways of using an assessment. For example, a measure that is valid for screening and identification may not be a suitable formative assessment. An assessment that is good for diagnosis may be a poor outcomes measure for program evaluation. Similarly, the validity of each of our 2019 assessment exemplars depends on how they are used in practice. Readers interested in exploring validity and reliability further can access the RAND Education Assessment Finder, which includes evidence of reliability and validity for specific SEL and academic assessments. Another aspect of technicality worth exploring and understanding is knowing which population scores are normed on, and whether that population is relevant to the student population one wishes to assess.

PRINCIPLE #6: ASSESSMENT IS PRACTICAL TO IMPLEMENT

Users find the assessment feasible and practical to implement.

Practical assessments are user-friendly and require reasonable investments of resources such as staff time and finances. Assessments are user-friendly if they can be quickly and easily learned by assessment administrators and other users. Training and technical assistance may be necessary and, if so, should be readily available and responsive to users' needs. Investments of staff time and finances are also important practical considerations for users and, therefore, should be within reach for decision-makers. Time and staff investments come into the picture at various stages of assessment implementation, including learning about the assessment, administering it, submitting scores, and receiving and interpreting data reports.

Decision-makers should carefully consider resource investments for implementing assessments, as they are often not straightforward. For example, administration time depends upon how the assessment is administered: Do students engage in the assessment together during class time or can they leave the room one at a time so instruction can continue? How long does the assessment take per student? How much staff time is required to support administration? Are these staff members who would otherwise be teaching? What are the constraints around score reports? For example, must scores be tallied by site staff? Are results automatically scored, analyzed, and reported in real time?

Here are two examples that illustrate these points.

- SELweb is simple to administer. Students taking SELweb can complete it relatively independently on a computer or tablet. Depending on how a school or district chooses to administer the assessment, students may be able to step out of a classroom to take the assessment, minimizing disruption to instruction. Scores are automatically calculated, taking the burden off school staff to interpret, and results are available immediately.

- Facilitators in the Afterschool Life Skills Programme complete pre- and post-assessments of participants' 5 life skills using the Dream a Dream Life Skills Assessment Scale. The time and resources needed to complete the assessment and use the results are built into the program, making assessment a feasible and practical component of the program.

---

8   Assessment Work Group. (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future.*

PRINCIPLE #7: ASSESSMENT DATA ARE USEFUL

The assessment provides data that is clearly useful for the intended purpose as well as potentially useful in multiple ways.

This principle lies at the crux of why many school districts use assessments at all: to make informed decisions about student learning. SEL assessment data, when used correctly (for the purposes for which the assessment was designed), can be a powerful contributor to decisions about student learning. A teacher may benefit from having access to SEL data on a daily basis to consult when making lesson plans, or, less often but still regularly, to check where students are in relation to benchmarks. A school or district may have specific needs for SEL data on an annual basis for academic year planning and goal setting, as well as at regular intervals throughout the year, and check progress toward goals. Students and their families may also wish for teachers to share data about the student at regular points in time, to inform their own decisions for supporting the student's development.

Practitioners should be clear about their assessment goals and use only assessments designed to help them meet those goals. This also requires that users recognize the value of an assessment system that does not rely too much upon any one assessment. That said, some assessment reporting systems are able to provide a variety of data reports that serve multiple users in an educational system. For example, they may report data by school, grade, classroom, student group, and/or individual student.

To be clear, no one assessment is suitable for the entire universe of assessment goals educators might contemplate. It is important to recognize that using any one assessment to achieve a large number of goals runs the risk that the assessment will be asked to do more than it is capable of, and that users will be tempted to make inferences and decisions based on data that are not appropriate for those purposes. Some questions practitioners might consider related to an assessment's usefulness are:

- When do we need the assessment data?

- What kind of information do we need to see in data reports?

- How do we want the data to be presented so it supports our decision-making processes?

Below are three examples of useful data, the first from a web-based assessment, the second from a site-developed assessment, and the third from program-embedded assessments. We focus on these because we were able to interview schools and programs that have demonstrated how they use data for multiple purposes at multiple levels. More detail about data use for each 2019 exemplar is available in Appendix D.

- SELweb helps districts identify trends across schools, support school- and classroom-level planning, and pair student buddies. SELweb provides score reports automatically and can show results at the individual, classroom, grade (within school), school/program, and district levels and includes scores for the individual domains measured as well as an overall SEL competence score.

- Schools within San Francisco Unified School District use SEL data for student report cards, to communicate with families, set schoolwide goals, make classroom decisions, and determine individual student improvement benchmarks.

- Two Rivers and NPDL use data from rubrics to inform decisions around how habits are taught, learned, and demonstrated in the classroom, as well as to understand districtwide trends and make plans for the coming year.

PRINCIPLE #8: ASSESSMENT IS A GOOD FIT OVERALL

The assessment is a fit between the needs and goals of the users and the qualities and properties of the assessment.

Assessment developers should also consider a final principle when designing direct assessments: the extent to which it is a good fit overall. This means, broadly speaking, that the assessment meets the needs of its user(s). More specifically, we encourage practitioners to consider the previous seven principles and carefully think through what each principle means in the context of their own educational setting. While there are many ways to consider "fit," it could look like a step-by-step focus on each principle resulting in a statement that defines the particular assessment goals under that principle. Once practitioners have each of these statements, they can then analyze them from a higher level to ask what a "good fit" overall means. The information and improved clarity that can result from such an exercise can help prioritize particular principles, set parameters to consider when gathering assessments, and help in the process of narrowing and selecting assessments to implement. Alternatively, for educational sites already using SEL assessments, this exercise can serve to reflect on how well the existing assessments are fitting current needs and answer questions about how to proceed.

As mentioned earlier, users may wish to consider particular aspects such as:

• Fit to intended purpose

• Fit to available budget

• Fit to time available

• Fit to data needs and existing decision-making processes

Conclusion

We set out in 2017 to explore what is happening with direct assessment and develop a set of design principles regarding the direct assessment of student SEL competence. While the field of direct assessment is still relatively young, there are now some promising examples of SEL measures and data informing decisions at multiple levels of an educational system across the globe. We hope these principles will challenge assessment developers who are ready to take on the challenge of designing direct assessments that meet these principles. We also hope these principles inspire a conversation around what else is needed to support the creation and effective use of direct assessments that drive social and emotional learning.

For those interested in learning more about design principles for assessment, please see the "Guiding Principles for Equitable Performance Assessment Systems," developed by the California Performance Assessment Collaborative. We have included these principles in Appendix F along with a link to where they are posted on the Learning Policy Institute website.

For those interested in learning more about SEL assessment, including direct assessment, we refer you to an informative report published in October 2019 by the Assessment Work Group titled _Student Social and Emotional Competence Assessment: The Current State of the Field and a Vision for Its Future_ as well as our 2017 and 2018 Design Challenge briefs.

About the Design Challenge

The Design Challenge is a three-year annual competition to identify design principles for direct assessment. By identifying and recognizing award-winning efforts, the Design Challenge seeks to stimulate the development and adoption of direct assessments of social-emotional (SE) competence supporting effective instruction and positive student development.

About the Assessment Work Group

The AWG is the working group for the Establishing Practical Social-Emotional Competence Assessments of Preschool to High School Students project, which was created to advance progress toward establishing practical SEL assessments that are scientifically sound, feasible to use, and actionable as a key priority for the field. The project's work group is a multidisciplinary collaborative of leading researchers and practitioners in the fields of PreK-12 education, assessment, social and emotional learning (SEL), and related fields. To learn more about the AWG, visit measuringsel.casel.org. To learn more about the Design Challenge, visit measuringsel.casel.org/design-challenge/.

Appendices

A. 2017 Design Challenge Exemplars

B. 2018 Design Challenge Exemplars

C. 2019 Design Challenge Exemplars

D. 2019 Design Challenge Exemplar Full Descriptions

E. Interim design principles from 2018 Revised Key Design Principles for Direct Assessments of SEL: Lessons Learned from the Second Design Challenge

F. Guiding Principles for Equitable Performance Assessment Systems from the California Performance Assessment Collaborative

Appendix A. 2017 Design Challenge Exemplars

Title	Skill Area(s) Assessed	Age Range	Description of Assessment
Student Assessment Engagement	• Self-regulation • Self-management	6th–10th grade	When students take an achievement test on a computer, metadata like the amount of time spent on each item are often collected. Research shows that students who often respond extremely fast–so quickly they could not have understood the item's content–are likely disengaged from the test. Our measure quantifies how often students respond extremely quickly over the course of a test, which is strongly correlated with scores from measures of social-emotional learning constructs like self-regulation and self-management.
Social Detective	• Perspective-taking	6th–12th grade	Panorama's Social Detective is designed to measure and help students practice social perspective-taking, a malleable and central social competency that underlies a vast range of social-emotional functioning at school and in life. In this performance task, students are challenged to be a "social detective" whose job is to figure out other people's values, interests, and perspectives. After watching short video interviews, students answer a series of questions to gauge how well they perceive and understand each person.
PERC	• Persistence • Effort • Resilience • Challenge-seeking	6th–10th grade	The PERC is a computer-based tool that assesses students' Persistence, Effort, Resilience, and Challenge-seeking behavior. These are key behavioral expressions of a growth mindset of intelligence. PERC is fully developed and available to potential users. It has been used by tens of thousands of students, for example, in evaluations of growth mindset interventions in South Africa, and in the Indonesian school system, as well as in hundreds of studies conducted on the Character Lab Research Network.
Zoo U Social-emotional Skills Assessment	• Communication • Cooperation • Emotion • Regulation • Empathy • Impulse control • Social initiation	3rd–5th grade	Zoo U provides a game platform for performance-based formative assessment of social-emotional skills in upper elementary grades.
The Calendar Task	Self-management, specifically • Organization • Time management • Prioritization • Planning • Scheduling	6th–12th grade	The calendar task (in development) will potentially provide a flexible and naturalistic platform to evaluate student self-management skills with the potential for multiple use cases, both summative and formative in nature.
PLUS Executive Functioning Assessment	Executive function skills as represented by • Inhibitory control • Working memory • Cognitive flexibility	PreK–12th grade	In order to employ direct assessments of executive function (EF) skills at scale, we developed a group-based assessment procedure that is time-efficient and cost-effective. We adapted four developmentally appropriate, widely used EF tasks for administration on tablet computers in a classroom setting. Our classroom protocol allows a minimally disruptive assessment of EF skills in all students at the same time.
Measuring Grit	• Grit	4th–12th grade	This assessment measures grit, i.e., the propensity to set ambitious goals, persevere in the face of failures, and put effort to build skill. We use an incentivized methodology that involves rewarding successful outcomes.

Appendix B. 2018 Design Challenge Exemplars

Title	Skill Area(s) Assessed	Age Range	Description of Assessment
Winners listed from First to Fifth Place			
Selected Response Assessment of Social-emotional Competence (SRASEC)	• Social knowledge • Behavioral knowledge • Emotional knowledge	K–5th grade	Computer game assessment that works to measure social-emotional competence and determine the relationships between direct and indirect measures of both social-emotional competence and academic achievement.
Text-Based Decision Game	• Discipline • Respect • Enthusiasm • Accountability • Maturity • Bravery • Initiative • Grit	9th–12th grade	Interactive text-based scenario game tailored to each student's academic career that measures a student's competency across a range of values.
Virtual Environment for Social Information Processing (VESIP)	Social Situations: • Ambiguous, • Provocation • Bullying • Compromise • Peer entry into a group • Friendship initiation	3rd–7th grade	Theory-based, web-based assessment that assess children's social information processing skills through social situation simulations. VESIP is fully developed and available. It is being implemented in 14 school districts and research partnerships, covering seven U.S. states and one province in Canada. A Spanish language version is being validated.
The Skills Rubric Template (SRT)	Spark Skills: • Communication • Collaboration • Problem • Solving • Innovation • Grit • Self-Management	9th–12th grade	Set of rubrics that assesses students' social and emotional skills. The SRT provides a singular definition for each skill and specific indicators to aid educators, coaches, and mentors to assign scores similarly.
Problem Solving Performance Assessment	• Utilizes SWOT analysis • Survey measures over 100 behavioral attributes	9th–12th grade	Internet-based survey for high school students that measures their level of proficiency in problem solving, examining their solutions with the SWOT Analysis.

Appendix C. 2019 Design Challenge Exemplars

Title	Skill Area(s) Assessed	Age Range	Description of Assessment
Web-based assessments that use virtual scenarios to assess performance			
Reflection Sciences Minnesota Executive Function Scale **(MEFS Application)**	3 Executive function skills • Working memory • Inhibitory control • Cognitive flexibility	Children as young as 24 months through adults	Game-based application, often using a tablet, completed by students with the help of an adult in approximately 5 minutes. Fluid readers and adults may use a self-administered version. Quick administration per child. Measures executive function skills starting at a young age. Sensitive enough to detect small changes in skill development. Accessible training and support. Curriculum strategies tied to results, for a fee. In use in over 200 locations, including 38 U.S. states and 18 countries. Available in 14 languages.
xSEL Labs SELweb Assessment	4 SEL competencies • Emotion recognition • Self-control • Social perspective-taking • Social problem solving	K-3th grade (SELweb EE) 4th–6th grade (SELweb LE)	Web-based direct assessment completed by students in approximately 30 minutes. Self-administered by even preliterate students. Easy administration and reporting. Provides real-time results. Identifies trends in SEL for district-level analysis. In use in over 200 schools in six U.S. cities.
Rubric-based assessments created within sites			
San Francisco Unified School District Standards Based Report Card	4 SEL competencies Social awareness Self-management Growth mindset Self-efficacy	Transitional Kindergarten through 5th grade	Observation rubric for teachers to assess students. Reported within report card. Connects with parents and families around students' SEL. Created by and in use in San Francisco Unified School District based upon CORE's Social-emotional Learning survey.
Two Rivers 5 Scholarly Habits	5 Scholarly Habits: • I know myself • I am independent and resilient • I show compassion and embrace diversity • I can connect and collaborate • I act with integrity	PreK 3–8th grade (3 years old to 14 years old) Broken down into grade bands: • Preschool–PreK (ages 3 and 4) • Kindergarten to 2nd grade • 3rd to 5th grade • 6th to 8th grade.	Observation rubric completed by teachers. Students also take self-evaluation and do short reflections typically at the end of the week. Developed by and in use in one charter school district in Washington, D.C. based on CASEL 5 competencies.

Rubric-based assessments created within programs			
Dream a Dream Life Skills Assessment Scale	5 Life Skills measures: • Interact with each other • Solve problems and overcome difficulties • Take initiative • Manage conflict • Understand and follow instruction	4th–9th graders (ages ranging from 8–15)	Multiple assessments embedded in afterschool arts and sports curriculum, including facilitator-conducted pre- and post-program observations of students (scale of 1-5) and 2-3 student stories of change for each grade (total of 12-18 stories per school) chosen and written by facilitators at conclusion of program for use by program. Use as part of the Dream a Dream After School Life Skills Programme (out-of-school-time program) Developed for underserved children in India. Six nonprofits in India use the assessment in their After School Life Skills Programme, expanding to Africa.
New Pedagogies for Deep Learning, Deep Learning Progressions	Deep Learning Progressions 6Cs • Communication • Critical thinking • Collaboration • Creativity • Citizenship • Character	Universal competencies implemented throughout NPDL in the K–12 age range	Measures of individual student development of the six deep learning competencies (aka the 6Cs) that can be used among students, teachers, and schools. Applicability to broad settings (schools, nonprofits, government) For districts interested in inquiry-based deep learning Supports learning from others via regional and global learning networks In use in eight countries and five U.S. states, 1,300 schools.

Appendix D. 2019 Design Challenge Exemplar Full Descriptions

Reflection Sciences Minnesota Executive Function Scale (MEFS Application)

https://reflectionsciences.com/

Reflection Sciences developed the Minnesota Executive Function Scale (MEFS App) assessment to measure the executive function skills: working memory, inhibitory control, and cognitive flexibility for children as young as 24 months. Their aim was to develop a more intuitive assessment that could provide accurate and necessary data identifying developmental delays in neurocognitive skills, known as executive function (EF skills, for nonprofits that work with at risk children and families. For example, they found that other assessments geared toward older audiences were not sensitive enough to pick up on all developmental variability (a 3-year-old receiving a 0 on an EF test may not mean they have 0 EF skills, rather the assessment wasn't sensitive enough to pick up on where the child is developmentally. According to Reflection Sciences, executive function skills serve to help one manage emotions, demonstrate empathy, and achieve academic success. The MEFS App is "played" using a game-based application, often on a tablet, and takes about five minutes for a student to complete with the aid of a classroom teaching assistant. To play the game, students must place animal cards into one of two boxes on the screen. For example, in level 5 of game play, a child may be asked to place all cards with green lions in the green lion box and all cards with orange cheetahs in the orange cheetah box for a couple rounds until the instructions change. As students progress to the next level of difficulty, the rounds for one type of instruction are shorter, causing kids to rapidly switch back and forth between various games and instructions within a level. This gradual switch and add-on instructional technique for each level of game play makes the assessment sensitive enough to identify where a child has difficulty following the rules of the game, leading it to indicate more precisely where they are at developmentally.

The MEFS App provides classroom level reporting and is used in more than 200 locations, including 38 U.S. states and 18 countries and available in 14 different languages. It takes about five minutes to complete either with help of a classroom teaching assistant aiding the student or through self-administration by fluid readers. Continuously working on improving its application, Reflection Sciences has expanded to include a flexible and user-friendly administration portal, and updated US national norms.

A District's Experience with the MEFS Application

To learn more about a district's use of the MEFS App, we spoke with staff at the St. Anthony-New Brighton School District. They administer the assessment three times a year: in September, to determine a baseline for students; in January, to evaluate students' executive function skills and make changes to curriculum; and in the Spring, to see how each student has grown since the beginning of the year and to gauge incoming students for kindergarten.

The district uses the application to measure student growth in their preschool classes. Kindergarten teachers use the data to learn where preschoolers are as they enter kindergarten. To get a more rounded view of the students' growth, they pair their January MEFS App results with data from the Teaching Strategies Gold assessment. If students are significantly behind their peers, the school begins to explore what supports they can provide to help the students. The data results have also helped teachers be more understanding of the developmental process that each student is at, and work toward helping the student along at their own pace.

The district has also found Reflection Science's professional development sessions crucial to their staff learning and classroom success. These sessions come at an additional cost and included EF curriculum and intervention strategies. One Reflection Sciences technique that St. Anthony-New Brighton's preschools has used when their preschoolers have traveled from one classroom to the next, is the "kitty exercise". In order to promote a quiet, single file line when traveling from one room to the next, teachers tell their students to imagine that they are each transporting a tiny scared kitty. They must whisper when they talk as to not scare their kitty, and must continuously pet their kitty to comfort it. This technique has proven to be extremely successful as it keeps their minds on the tasks at hand and helps them practice mindfulness, self-control, working memory, and inhibitory control in the process.

Teachers can access scores immediately in the MEFS App cloud application. The raw data scores of the assessment are converted to standard scores using a national sample of same-aged students as the reference group, recently being re-normed on 32,800 children in the U.S. ages 2-18, as well as 1,340 adults. Individual and school-level reports currently are available.

SELweb Assessments

https://xsel-labs.com/

xSEL Labs developed SELweb assessments of social and emotional competence for elementary-aged students. SELweb is a web-based performance task measure that assesses students' emotion recognition, self-control, social perspective-taking, and social problem solving. SELweb has been used thus far in the United States in over 72 schools across six cities to measure SEL outcomes.

SELweb is comprised of two grade band assessments, one for Earlier Elementary (EE) for children in grades K-3, and the other for Later Elementary (LE) for children in grades 4-6. The assessment takes 25 to 30 minutes. It has several performance task segments, including:
- Facial recognition: student identifies how a person feels based on their facial expression
- Self-control: a click and drag task that has a delayed cursor response
- Scenario on social perspective taking: how does a person feel in that moment?
- Scenario on social problem solving: how would you react if...?

In the scenario segments, the child must determine how they should react. For example, the child must determine if another student is being mean on purpose and if so, would they (a) notify an adult, (b) be mean back, or (c) ignore them.

In order to provide rich data to its educators, xSEL Lab focuses on providing their assessments to networks of schools within a district and whole school districts. By doing so, they are able to provide districts with significant amounts of data from which they can track school level areas of improvement over time as different developmental approaches are implemented. A key aspect to tracking this data progression is administering the assessment twice a year, once in the Fall and again in the Spring.

Reports are available to users with permission, including students, and can be modified to show individual, classroom, grade-level within school, and school district results. Reports show results for each of the 4 competencies measured and an overall SEL competence score (the raw assessment scores are converted to standard scores using a national sample of same-aged students as the reference group). Student data also includes performance indicators: above, meets, below or well below expectations in each area. Reports can also show change over time.

A School's Experience with SELweb

To learn more about the school site level use of SELWeb, we spoke with staff at William F. Murphy Elementary and Scott School in Illinois. Before taking the assessment, students learn of the importance of growing their whole selves and taking pretests of their emotional skills, just as they do for math and other subjects. From the fall baseline data, teachers and staff are able to see the strengths and needs of their students and classrooms. Teacher groups sit down with administration and instruction coaches to discuss gaps and areas of concern. As a method of best practice response, teachers are also mixing their classrooms and pairing students with various SEL proficiency skill levels on projects, at recess, or even at lunch so that students can also grow in SEL knowledge from their peers. Coaches offer alignment guides with the Second Step SEL program. Second grade teachers also use the data to pair students who aren't proficient in SEL with those that are as a part of a buddy program across the three second grade classes.

San Francisco Unified School District Standards-Based Report Card

https://archive.sfusd.edu/en/curriculum-standards/report-cards.html

As a member of the CORE Districts of California, San Francisco Unified School District (SFUSD) has used CORE's Social-emotional Learning survey since 2014. To incorporate SEL more fully into their district, SFUSD created the Standards-Based Report Card (SBRC) for Transitional Kindergarten through 5th grade. The SBRC is a way to communicate to parents about their child's progression to meet standards. The report card includes SEL assessment scores in addition to academic scores.

The SEL assessment is an observational rubric that reports students' social awareness, self-management, growth mindset, and self-efficacy. Teachers grade students on a four-point scale to allow for more precise student ratings with students receiving a report card three times a year. Though the SEL assessment has been incorporated in the district's report card since 2014, these specific four skills were formalized in the 2016-2017 school year. Alongside the assessment, SFUSD created a teacher guide to help their teachers create safe and supportive classroom environments through the four SEL skills, as well as an observational grading rubric to help them identify students' skills along the four-point scale.

Approximately 30 schools in SFUSD are grant funded to explicitly teach SEL in the classroom and have SEL support from teachers on special assignment. The grant allows 10 new schools each year to pair SEL curriculum with the SEL assessment. The school chooses the curriculum, with the most common ones being the Second Step SEL program and Kimochis.

At each school, leadership teams are created to support teachers and help with school-wide goal setting for the academic year. The leadership teams are comprised of the principal, social worker, a representative from the Culture Climate team, special education staff, and some teachers. Each school site also has a Teacher on Special Assignment with multiple roles, one of which is to collect school-level data on SEL and Culture/Climate (e.g., school safety, sense of belonging, and other related content). They send the data to SFUSD's Early Education, SEL, and Literacy Initiatives department for analysis. The analyzed data is then used to consult with the leadership team on schoolwide goal setting for the following year.

Teachers on Special Assignment also directly support teachers. For example, at the beginning of the year, they review the scaling rubric data together, as well as school site data from the previous year, and begin to strategize around best practices for each SEL skill, reflect on what they currently do well in the classroom and what they would like to try. The pair then go over school reports to see the growth of incoming students.

Teachers can use the SEL data to set student improvement benchmarks and classroom decisions. For example, students approaching a two on the four point scale ("with moderate teacher supports, student can..."), a teacher could set an improvement benchmark of a three. Some teachers have used the data in one-on-one conferencing with the students or to communicate with families at parent teacher meetings and provide parents and guardians translated family guides containing tools to support SEL development at home.

Two Rivers 5 Scholarly Habits

https://www.tworiverspcs.org/

Two Rivers Public Charter School District developed the 5 Scholarly Habits assessment of social and emotional competence for all its students (ages 3 to 14) as part of a district-wide SEL/culture initiative. The 5 Scholarly Habits are an observational rubric that teachers fill out to measure students' competencies according to the following statements: I know myself; I am independent and resilient; I show compassion and embrace diversity; I can connect and collaborate; and I act with integrity. Students receive classroom instruction on the five habits and are graded on them twice a year in addition to completing self-evaluations and short reflections on their performance in one skill area at the end of the week. Using common language across all grade levels and classrooms, these habits are embedded in the districts' culture from classroom instruction and environment, small student group mentoring, recognition swag, to weekly school assembly shout outs. Two Rivers Public Charter School District's two elementary schools (PreK age 3 to 5th grade) and one middle school (6th to 8th grade) serve a total of 850 students in the Washington, D.C. area.

The observational rubric, which Two Rivers refers to as its SEL rubric, contains standards on a 1-5-point scale for each scholarly habit. Each standard in turn has corresponding "look fors" with a rating range of 1, 3, 5 that helps teachers identify to students what they are regularly looking for in class. The standards and "look fors" were developed by teachers who worked together in grade-level bands to ensure that the habits are consistent and recognizable to children as they receive instruction on them throughout the year and as they grow up and change grades. Overall, their rubric breakdown from measure, to standard rating, to "look for" rating, has established clearly identified grading guidelines that helps their district collect more precise data.

Grade-level bands are broken into the following groups according to similarities in students' cognitive abilities and classroom routines: Preschool and PreK (ages 3 and 4); kindergarten to 2nd grade; 3rd to 5th grade; 6th to 8th grade. Teachers and staff intentionally worked together across grade level bands to create SEL language and teaching dialogue that would be valued and understood by all students. Consistent teacher dialogue is crucial to help students (especially younger students) understand and learn the 5 scholarly habits. To teach "I show compassion and embrace diversity" teachers observed what was going on in their classrooms and identify situations that they could use as teachable moments. For example, teachers saw that students were playing with the same kids. So, they talked with their classes about how nice it is to hang out with people that don't look like [them] because you get different perspectives and learn new games. Phrasing the measure in this way (specifically learning new games) showed big results. As teachers see good scholarly habits in action, they also acknowledge what the student is doing so that the student connects that type of behavior to that scholarly habit. For example, a teacher might make a connecting acknowledgment by saying: "You just showed me that you're being compassionate because you helped a friend who was upset." In teaching the scholarly habits, teachers have noticed students voluntarily going up to one another to help their fellow classmates, a decline in behavioral issues, overall positive behavioral changes, and closer relationships between students and teachers. To ensure the stability of their SEL system they have two teachers in every class (one main teacher and one assistant). Thus, even when a substitute is present the SEL skills are still being carried out by another classroom instructor.

The data generated from using the SEL rubric helps inform decisions around how habits are taught, learned, and demonstrated in the classroom, as well as to understand district-wide trends and plan for the coming year. For example:

- Teachers can review data for their "crews" (mentoring student group sessions), classrooms, and grades, and collaborate to make content decisions and create intentional opportunities for students to continue progressing along the standards for the 5 habits.

- Grade-level teams and crew leaders meet once a week to focus on a habit they want to work on with students that has been flagged from the behavior data and grade and assessment data. During this time, they also talk about how they can be more intentional about providing opportunities for students to set them up for success and meet expectations the school has set forth. Teachers can use the data to identify where mentor and other SEL supports may be needed to help students achieve their aspirations as they move from grade to grade.

- Teachers and staff revisit the data at the end of the school year to look for trends across all grades and determine the focus and content standards for the coming academic year.

- Through TR (Two Rivers) Connect, staff conducts quarterly check ins with former 8th graders so see how they are doing and reflect on how Two Rivers helped them.

Dream Life Skills Assessment Scale

https://dreamadream.org/

Aimed at serving children from vulnerable backgrounds, Dream a Dream developed their Dream Life Skills Assessment Scale for 4th-9th graders. The Life Skills Assessment Scale is an observational assessment that looks at the following 5 Life Skills measures: interacting with each other, solving problems and overcoming difficulties, taking initiative, managing conflict, and understanding instructions. Dream a Dream views these skills as interdependent, where the presence of one skill suggests the presence of the other Life Skills.

The Dream Life Skills Assessment Scale originated in India in 2002 and is an open source tool that is currently used in 25 schools and two community centers in Bangalore and at least six nonprofits across six Indian cities. Globally, the assessment is known to be used in at least four nonprofits across four countries with plans to pilot the scale in various African cities next year.

The Life Skills Assessment Scale is a part of Dream a Dream's art- and sports-focused After School Life Skills Programme, which is offered in partnership with various schools. Students in grades 4-9 (ages ranging from 8-15) decide if they want to participate in either the art or sport SEL centered program. For example, kids participate in athletic activities that require teamwork, art activities where they create skits, write or direct, or work together in groups to create a single painting. In each activity, children are challenged to work on the 5 Life Skills. Each grade level's afterschool program lasts for two hours and takes place once a week Monday through Saturday beginning in June and ending in March (corresponding to their yearly school schedule). The afterschool programs for ages 8-15 are conducted by 40 Dream a Dream facilitators, 26 of which are graduates of schools that provided these programs.

To gauge students' life skills, pre- and post-assessments are conducted along with daily session reports. For pre-assessment, the first three program sessions are used for students to get comfortable with the program. By the fourth session, students are observed by the facilitator as they've had time to adjust to the sessions and are more likely to be exhibiting their actual development skill level. After the fourth session pre-assessment, facilitators conduct daily session reports and submit them to Dream a Dream's Monitoring and Evaluation team to analyze. Once baseline data is submitted, it is then reported back to facilitators in multiple ways, including: gender, program (e.g., sports or art), life skill, individual child data, batch data, and school-wide data that is also compared to developmental norms so facilitators know where each child is at and can design specific interventions for them throughout the year. This data also helps validate the facilitator's observations. Correlations are also drawn with attendance, longevity of participation in a programme, and retention data. At the end of their academic year, facilitators then submit end of the year observational assessments for each student, rating their improvement on a scale of 1-5, in addition to 2-3 student stories of change for each grade (total of 12-18 stories per school) These stories are used to view the progress, effectiveness, and impact of the assessment and program, for further research by other NGOs, and to update funders. The Life Skills Assessment Scale work is also rooted in Dream a Dream's Career Connect (ages 15-20) workshops that prepare older students for the workforce by teaching them about technology, internships, money management, broken English, as well as their Teacher Development programs where they build teacher capacity around nurturing child development of the 5 Life Skills and respond to students in a more empathic way.

Deep Learning Progressions Assessment

https://www.npdl.global/

The **Deep Learning Progressions,** developed by New Pedagogies for Deeper Learning (NPDL), is used by teachers to measure each student's developmental progression of the 6 Deep Learning Competencies, or 6Cs: communication, critical thinking, collaboration, creativity, citizenship, and character. There are 6 Deep Learning Progressions, one for each competency, that highlight various dimensions of the showcased competency and displays a five-point rating criteria describing what a student needs to exhibit in order to receive that grade point. The five-point scale ranges from: limited evidence, emerging, developing, accelerating, to proficient. Teachers use the Deep Learning Progressions as direct measures of students' development of each competency. To aid their grading, assessment evidence such as assignments, projects, tests, conversations, observations are used to map student ratings. With a focus on system-level change, NPDL offers tools to support teachers and system leaders. The teacher tools include: design protocol—help to design deep learning classroom experiences, design rubric—measures effectiveness of their deep learning experience design, and teacher self-assessment—assess their ability to design deep learning experiences for their students. NPDL also offers three conditions rubrics to help principals and leaders measure their school's, cluster's (district or multiple schools), and system's (state/country) status and ability for implementing deep learning.

Schools use their school-site data to determine how many (and which) of the 6Cs they would like to focus on as a community for the year (most choose 1-2). To gauge where the school is at and what is needed to improve, teachers from each grade may pull pre- and post-self-assessment data from (roughly six) students around the school's chosen Cs to share with the rest of the staff. At these staff meetings, teachers also bring their teacher design rubric and self-assessments to discuss best methods of practice and brainstorm deep learning improvements for their classrooms. To promote self-sustainability NPDL provides coaching to teachers and leaders to help them understand the work, how to reassess their design protocols, and see themselves in the 6Cs. They then share their learnings and how-to knowledge with colleagues for further scaling.

The Deep Learning Progressions have been used in eight countries and five U.S. states and works globally with school clusters (100+ schools) and networks (5+ schools) to build knowledge and practices that develop deep learning and foster whole system change. NPDL provides various opportunities for in-person and digital collaborative learning through its Deep Learning Hub: schools, districts, and countries are connected digitally and can attend both regional and global convenings to learn from their peers and share their findings.

Overall, NPDL has noticed significant mindset shifts from both staff and students that use their rubrics. With teacher turnover and student misdemeanors dropping, and student attendance and engagement increasing across the board, the data reveals that both groups are not only recognizing more value in themselves, but in the work they are doing.

A District's Experience with the Deep Learning Progressions

To learn more about a district's use of the Deep Learning Progressions, we spoke with staff at Burlington-Edison School District. There, students in kindergarten through 2nd grade use a student version of the Deep Learning Progressions to rate themselves on a smiley-to-frowny face scale, while 3rd through 8th grade students write down their reflections. Teachers model and share examples for the younger grades to help them understand how to rate themselves.

Though the schools have students work toward developing in all 6Cs, Allen Elementary (a school in Burlington-Edison School District) chose for their school to emphasis on Community and Collaboration. This decision was made by their leadership team composed of teachers, instructional assistants, special education staff, administration, and a representative from each grade band (each grade has its own band, except 7th and 8th which are combined). As a result, teachers from each grade band decided to emphasize either Community or Collaboration as their classroom focus for the year, with each teacher deciding how they would implement it. In 7th-grade English, students analyzed literary characters based on two of the 6Cs and tracked the character's growth along the Deep Learning Progressions. They would then state what they did to identify the character's placement along the Deep Learning Progression scale, as well as identify textual examples of the character's growth.

Similar to NPDL's findings, the district has noticed a shift in teacher mindsets. They're asking themselves what each student's resilience is when they tackle a tough problem and are starting to recognize that students have other talents. Middle School grades in the district used to include 0s for incomplete assignments and teachers are now considering what that really means and, in some cases, trying to eliminate 0s. By also asking students what they really need, discussions around eliminating homework and making sure students are not negatively impacted because their home life can't support a homework schedule have also arose.

Appendix E. Interim design principles from 2018 Revised Key Design Principles for Direct Assessments of SEL: Lessons Learned from the Second Design Challenge

After the first Design Challenge, we established principles to guide future SE competence direct assessment development. Among those principles were:

- Usability and feasibility in authentic educational settings.
- Clarity about the purpose for which the assessment is designed.
- Developmental and cultural appropriateness.
- Data reports that are easily understood and address the specific information needs of educators.
- Feasibly and easily used by entire districts with minimal intrusion on instructional time.
- Technical soundness.

Those design principles remain relevant to the ongoing development of SE competence direct assessments.

First, assessments that are most responsive to practitioner needs appear to be those that provide opportunities for students to demonstrate their SE competencies in response to performance challenges encountered in the real world. This suggests that simulations or vignettes may more closely align to practitioner needs than tasks that may have technical merit but are decontextualized or do not have clear and obvious application in real-world interactions.

Second, involving practitioners in the process of assessment development will strengthen the products of these efforts. We learned directly about the benefits of practitioner involvement from our two-stage Design Challenge, in which we solicited practitioner views of the greatest assessment needs followed by a call for submissions of direct assessment designs, prototypes, or fully developed assessments. In our second year, more practitioners submitted designs, and those received had high face validity. It was clear what the assessments were designed to assess, and it was easy to imagine how they would be used to guide practice. This suggests that partnerships between practitioners and test developers that start at the very beginning of assessment development—identifying the assessment need—will yield better assessments than would be possible without practitioner involvement.

Third, direct assessment is not the only strategy that can meet practitioner needs. A careful consideration of the needs described by practitioner submissions suggests that a variety of assessment methods might be useful for addressing their needs. For example, one practitioner expressed a wish for an assessment to evaluate skill generalization across contexts. While direct assessment might be helpful, frequent self-report or "experience sampling" might also be an effective method. Another practitioner wished to measure parent and child understanding of social-emotional skills after instruction. Direct assessment of those skills might be useful, but simple self-report to assess knowledge might also serve that goal. There is a strong need for direct assessment, and therefore it has been the focus of the AWG's Design Challenge. However, because the varied goals of SE competence assessment may not all be best served with direct assessment, it may be useful to consider the advantages of offering a broad portfolio of assessment options that can be used alone or in combination to provide a clear picture of student SE strengths and needs.

Appendix F. CALIFORNIA PERFORMANCE ASSESSMENT COLLABORATIVE Guiding Principles for Equitable Performance Assessment Systems

1. Features of the performance system are aligned to clearly articulated student competencies focused on assessing readiness for postsecondary success in college, career, and civic life.

2. The system is designed to take into account the needs of the most underserved populations (such as English language learners, recent immigrants, and students with disabilities) so that they can demonstrate what they know and are able to do.

3. The system includes the collection and/or exhibition of evidence of student growth and proficiency in relation to the competencies.

4. Presentation before an authentic audience allows for students to demonstrate their knowledge and skills by sharing their learning publicly.

5. Students have multiple opportunities to develop and demonstrate mastery: if their exhibition or portfolio is not proficient, there is an expectation of revision until proficiency is achieved.

6. Performance is evaluated according to rubrics and/or clear criteria of competence across multiple dimensions of performance.

7. Outcomes from the performance assessments provide data on how students are performing in order to improve curriculum and instruction.

8. The system enables students to take ownership over their own learning and growth and allows them to make choices about what they develop and exhibit.

9. Reflection is a critical component of the performance assessment process. Students engage in metacognitive processes to reflect on, assess, and improve their own work and to plan, with their teachers and peers, for future learning and growth.

10. The development of students' social-emotional skills, such as growth mindset and resilience, is included throughout the performance assessment process.

Source: https://learningpolicyinstitute.org/sites/default/files/CPAC_Guiding_Principles_2018.pdf