

The logo for BUROS (Center for Testing Assessment Literacy) is positioned on the left side of the cover. It features the word "BUROS" in a large, bold, white sans-serif font. Below it, the words "CENTER FOR TESTING" and "Assessment Literacy" are written in a smaller, white sans-serif font. The logo is set against a dark blue background with a large, light gray circular graphic element that overlaps the text.

**BUROS**  
CENTER FOR TESTING  
Assessment Literacy

# **Social-Emotional Learning (SEL) Assessment Technical Guidebook**

This guidebook maybe be copied and reproduced for educational and research purposes. When disseminating cite Buros Center for Testing and Spencer Foundation's role in this work.

Buros Center for Testing–Spencer Foundation Project Scholars (2020). *Social-Emotional Learning Assessment Technical Guidebook*. <https://buros.org/sel-assessment-technical-guidebook>

# Table of Contents

<u>About the Project and Guidebook</u>	1
<u>Buros-Spencer Project Scholars</u>	2
<u>Introduction to SEL Assessment Technical Guidebook</u>	3
<u>Guide 1: Locating and Identifying a SEL Assessment</u>	5
<u>Guide 2: Evaluating the Measurement Quality of a SEL Assessment</u>	10
<u>Guide 3: Recommendations when Using Social and Emotional Learning (SEL) Assessments</u>	12
<u>Appendices</u>	16

## About the Project and the Guidebook

In the past decade, there has been an increasing interest in the measurement of students' social and emotional learning (SEL) for decision-making purposes in various educational contexts. Federal and state educational policies indicate trends in basing accountability and improvement for student learning on more than just cognitive skills. There is also an increasing emphasis on the use of these skills in college admissions decisions in addition to K-12 learning.

The use of non-cognitive measures has prompted concerns regarding the validity, reliability, and fairness particularly when making high-stakes decisions. Professional testing standards exist that generally address the need for evidence to support these elements of an assessment. However, SEL assessments pose unique challenges due to different types of validity threats and potential fairness issues that accompany them.

Therefore, the primary goal of this project led by the Buros Center for Testing and funded by the Spencer Foundation was to bring together the most informed psychometric scholars in SEL research to synthesize what is currently known and unknown about the validity, reliability, and fairness of these measures for primarily educational uses. This knowledge was then applied to develop a list of guidelines for educators as to the most important considerations to make before selecting or using SEL assessments for decision-making purposes.

Questions about this SEL Assessment Technical Guidebook can be directed to Dr. Jessica L. Jonson, [jjonson@buros.org](mailto:jjonson@buros.org).

## Buros-Spencer Project Scholars

**Jeremy Burrus**

ACT

**Susanne Denham**

George Mason University

**Steve Elliott**

Arizona State University

**Kurt F. Geisinger**

Buros Center for Testing  
University of Nebraska–Lincoln

**Laura S. Hamilton**

RAND Corporation

**Thomas P. Hogan**

Scranton University

**Jessica L. Jonson**

Buros Center for Testing  
University of Nebraska–Lincoln

**Randy Kamphaus**

University of Oregon

**Michael Rodriguez**

University of Minnesota

**Neal Schmitt**

Michigan State University

**Beverly J. Vandiver**

Western Michigan University

**Frank C. Worrell**

University of California – Berkeley

### Additional Contributors

**Clark McKown**

Rush University Medical Center

**Pam Trantham**

Cambium Assessment

# Introduction to the Social and Emotional Learning Assessment Technical Guidebook

Assessments can provide educators with information about students' status and growth in desired competencies. This information can then be used with internal and external constituents for both accountability and improvement. In the case of social and emotional learning (SEL) where policies and practices are dynamically evolving, many PreK-12 educators struggle to select technically sound assessments that can appropriately guide practices and contribute to valid interpretations of student learning.

## Five Reasons Why Selecting and Using SEL Assessments Can Be Difficult

### **Reason 1: Lack of consensus about SEL definitions**

Consensus definitions of SEL are not widely recognized; as such, there are several frameworks that advance a variety of competencies. Definitions and frameworks play central roles in prioritizing skills that assessments measure and educators teach.

### **Reason 2: A focus on SEL strengths instead of problems**

Traditionally assessments of social and emotional functioning have focused on problem behaviors, rather than social and emotional strengths children should know and be able to demonstrate. Current SEL competency frameworks emphasize more of the latter strengths than the former problems.

### **Reason 3: Alignment between SEL assessments and the purpose for assessing SEL**

Technically sound assessments are needed for different purposes and few have been designed for all of these purposes. These purposes include identifying students' SEL skill development needs and making decisions about student and program outcomes. In particular, the measurement quality of many SEL assessments has not been substantiated with comprehensive methods and representative samples of students.

### **Reason 4: Use of a wide array of informants and methods in SEL assessment**

SEL assessments collectively use different types of respondents (e.g., students themselves, teachers, parents, and peers) and various formats (e.g., behavior rating scales, problem-solving situation tests, knowledge tests). This array of respondents and formats results in assessments that vary on important dimensions such as administration time, content coverage, cost, and threats to the validity of score inferences.

### **Reason 5: Accessibility and fairness of SEL assessments**

SEL assessments, like all assessments used to make important decisions that influence students' education, must be accessible and fair. To date, most SEL assessments have not demonstrated through rigorous research that they are both accessible and fair for students of different genders, ages, cultures, and linguistic backgrounds.

For these reasons, the Spencer Foundation funded a project led by the Buros Center for Testing to develop a user-friendly technical guidebook for educators involved in selecting and overseeing the use of SEL assessments with students in PreK through Grade 12. The result are the following three guides:

### **Guide 1: Locating and Identifying a SEL Assessment**

### **Guide 2: Evaluating the Measurement Quality of a SEL Assessment**

### **Guide 3: Recommendations when Using SEL Assessments**

We encourage educators to review, share, and apply these guides when engaging in the selection and use of SEL assessment.

# Guide 1: Identifying a Social and Emotional Learning (SEL) Assessment

Educators are more likely to choose a better product if the appropriateness and relevance of different options are examined and evaluated systematically. Therefore, this guide describes a systematic process for locating and identifying a social and emotional learning (SEL) assessment that aligns with the needs of the educational program and the setting in which the assessment will be used. This process is ideally completed with a group of individuals who have knowledge of SEL programs, expertise in assessment/measurement, and who will use and interpret SEL assessment results.

When locating and identifying potential SEL assessments the process should include the following three steps:

Step 1: Clarifying the Purpose for SEL Assessment

Step 2: Identifying Appropriate SEL Assessments

Step 3: Review Technical Evaluations of SEL Assessments

## Step 1: Clarifying the Purpose for SEL Assessment

### What SEL competencies are you interested in measuring?

Specify and clearly define SEL competencies so that decisions about what assessment to use are consistent with the intended competencies.

SEL frameworks help educators understand how competencies are defined; how they develop in students over time; and how they link to standards, curricula, and instructional tools. For more information about SEL frameworks, refer to [Adopt an SEL framework](#) in CASEL's AWG SEL Assessment Guide.

### What specific decisions will be informed by the SEL assessment results?

Assessment occurs in order to better inform various uses, and there are a broad range of intended uses that can be informed after assessing SEL, such as determining students' strengths and needs, evaluating an SEL program, tracking SEL over time, and improving school quality. Not every assessment is designed for every intended use, so by articulating how SEL assessment results will inform intended uses, you will be better able to evaluate and select an assessment that was intended for that purpose.

Ideally, SEL assessment should be implemented along with a conceptual model that links the implementation of SEL programs or practices to specific short- and



long-term learning goals. For more information about conceptual models, refer to [Develop a theory of change](#) in CASEL's AWG SEL Assessment Guide.

### **For what students and in what school setting will SEL assessment be used?**

Assessments are designed for a particular group of students in a particular setting and then developed with representative sample(s) of those students and in those settings. Selecting an assessment that aligns with the students and the setting in which the SEL assessment will be used is important for valid interpretation of scores. A few example considerations include school type (public, private, charter); ages or grades of students; and student demographics, particularly in regard to ethnicity/culture, multilingual learners, and students with disabilities.

### **How will SEL assessment be administered and scored?**

Consider early in the process who will be responding to the SEL assessment and who will be administering and scoring the assessment. Some assessments may require formal training or specialized practice for administration, and failure to use specially trained administrators could lead to improper administration/scoring, interpretation of results, and invalid decision-making. In addition, different assessment formats have strengths and limitations that should be balanced with factors such as time, cost, in-depth assessment of narrow areas, and less in-depth assessment of broad areas. The assessment you select may require compromises, so it is helpful to think about these early in the process.

Typical forms for SEL assessment include self-rating scales, observation or interview protocols, and/or performance assessments. For a detailed listing of the strengths and limitations of different assessment formats, refer to [Step 4: Review the Assessment Options](#) in CASEL's AWG SEL Assessment Guide.

## **Step 2: Identifying Appropriate Assessments**

Keeping in mind the purposes clarified above, you can now attempt to identify SEL assessments that might be appropriate for those purposes. Below are several sources that could be used to create a list of one or more SEL assessments that appear relevant for your purposes and are candidates for further inquiry into technical quality of the measure prior to selection. These include both specific sources for identifying SEL assessments and more general, comprehensive sources.

## **SEL-specific sources for identifying assessments**

Three sources provide lists of assessments specifically related to SEL. These sources can be helpful in your initial search for an appropriate measure, but the identification alone is not an endorsement of its match with your local purposes or the technical quality of the measure.

[The Assessment Catalog](#) in CASEL's AWG SEL Assessment Guide outlines steps in selecting an SEL assessment and provides a list of more than two dozen SEL assessments. Each listing offers basic information about such things as target constructs and age groups, administration times, and links to publisher information about the assessment.

[The RAND Education Assessment Finder](#) is a web-based, searchable tool with information for approximately 200 assessments of interpersonal and intrapersonal competencies as well as cognitive abilities. Each assessment listing includes basic information such as a brief summary of validity and reliability evidence, and populations from which available technical evidence has been collected.

[The American Institutes for Research \(AIR\)](#) offers a Tools Index with a list of several dozen SEL assessments organized by age/grade level that provides basic information such as target constructs, time for completion, and authors. The site indicates plans to update the list annually and that several assessments are available but not sold commercially.

## **Comprehensive sources for identifying assessments**

There are also four more comprehensive sources that include SEL assessments within a broader listing of assessments.

[The Test Collection at ETS](#) is a freely available, searchable online database with entries for over 25,000 assessments that provide basic information such as target constructs or traits, age or grade ranges, source (e.g., publisher or journal article), and such matters as number of items and time for administration. Not all assessments listed are commercially available but information on how to obtain specific assessments is provided. Note that some entries can be dated given that the listing is cumulative.

[Tests in Print \(TIP\)](#) is a reference volume published by the Buros Center for Testing that provides descriptions of more than 2,300 commercially available tests in English. Each listing details such information as test purpose, scores, population, administration time, publisher, and price. TIP is available in hard copy as well as

electronically via the searchable EBSCO or Ovid databases in most academic libraries across the country or through the Buros website [Test Reviews Online](#)

[Pruebas Publicadas en Español \(PPE\)](#), a reference volume also published by the Buros Center for Testing, is the Spanish counterpart to TIP in that it offers descriptive listings of more than 600 commercially available tests offered either wholly or partly in Spanish. Listings are given in both Spanish and English and provide the same information as TIP as well as information about the test's origin, norms, translation/ adaptation process, and test components available in English and Spanish. PPE is sold in both hard copy and electronic (PDF) formats and may also be accessed online through EBSCO/Ovid databases at subscribing libraries.

[PsycTESTS](#) is a searchable electronic database available as a subscription or through academic libraries sponsored by the American Psychological Association. It contains approximately 55,000 entries. At least 75% of those listings are available but not sold commercially. Listings provide purpose, constructs, administration information as well as minimal information on reliability, validity, and norms. For some entries, the assessment itself is attached, along with an indication of whether its use requires special permission.

### **Alternate sources for identifying assessments**

Professional journals feature articles, or collections of articles in special issues, oriented toward SEL assessments. Some articles will provide listings and brief descriptions of the assessments and other times more evaluative summaries. Such articles, when recent, can be very helpful, but their appearance is quite unpredictable.

Other users in your professional circles may be a good source for identifying SEL assessments overlooked in the other sources. However, keep in mind that just because a particular SEL assessment is appropriate for one setting does not necessarily mean it is appropriate for your own setting without further evaluation.

## **Step 3: Review the Technical Evaluations of SEL Assessments**

With the identification of a subset of potential SEL assessments that align with your purposes, the next step would be to attempt to obtain existing technical evaluations for those SEL assessments. There are several sources in which technical evaluation typically by qualified professionals can be found.

## **Mental Measurements Yearbook**

[The Mental Measurements Yearbook \(MMY\)](#) series by the Buros Center for Testing publishes professional reviews of commercially available tests by two qualified professionals. Reviews include a description of the test, development procedures, technical characteristics (reliability, validity, fairness), and a summary commentary on the quality of the test. The most recent edition (the 20th MMY) contains reviews for about 200 tests and is available in hard copy and electronic database form in most academic libraries. Reviews of individual tests may be purchased directly from the Buros Center for Testing's [Test Reviews Online](#) website.

## **Professional Journals**

Professional journals, as noted earlier, provide articles that can sometimes assist in identifying SEL assessments as well as provide reviews of the technical quality of the assessments. Keep in mind that authors of the articles may not always be independent in terms of conflict of interests (e.g. author is the assessment developer).

## **Technical Manuals**

Assessment publishers' technical manuals are the primary source of technical evidence for an assessment, but that evidence may not always align with professional standards and guidelines. Information in manuals can be technical and difficult to read without the help of someone with psychometric expertise. However, an assessment's technical manual is the primary source for conducting a technical evaluation; any assessment should have a manual or report available that provides detailed information about the assessment's purpose, development, structure, norms, reliability, validity, and fairness, as well as administration and scoring procedures.

## **Professional Consultants**

Professional consultants may help evaluate the technical quality of the assessments you are considering. Such persons might come from local universities or research centers but should not have a conflict of interest in SEL assessment.

## Guide 2: Evaluating the Measurement Quality of Social and Emotional Learning (SEL) Assessments

Once potential SEL assessments have been identified using the three steps in Guide 1, educators should conduct their own evaluation of the measurement (technical) quality of a social and emotional learning (SEL) assessment by identifying what types of information and evidence a developer has provided for an assessment and determining from that information if the assessment is appropriate for local plans for interpretation and use. Technical (psychometric) considerations addressed in this guide are based on the [2014 Standards for Educational and Psychological Testing published by the American Educational Research Association \(AERA\)](#), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

### **The purpose of this guide.**

The purpose of this guide is to prepare assessment users to know what questions to ask regarding the availability of information and empirical evidence that may support the intended interpretation and use of an SEL assessment for their student population. The extent to which the technical evidence for an SEL assessment addresses these questions will assist informed selection decisions that will translate into more valid interpretations and uses.

### **The intended users of this guide.**

This guide is for educators tasked with the selection and use of an SEL assessment. Ideally, a group of educators and experts with relevant insights about the content to be assessed and the use of the resulting information to guide action will be involved in the evaluation. Although specialized psychometric expertise can be helpful, this guide was written for assessment users who may not be experts in the technical details of assessment development.

### **Types of interpretations and uses addressed by the SEL assessment guide.**

This guide is applicable to situations where SEL is measured to provide feedback and improve instruction and programs. The guide is not intended for situations where SEL assessments are being used for accountability or in consequential decision making at a group or individual level.

Consequential decisions at an individual student level would involve measuring student learning to screen or diagnose students in need of additional services or intervention or to identify students with a mental health concern. If an SEL

assessment will be used for these types of consequential decisions, educators should consult with school district professionals who have received appropriate training and who hold the licenses or certifications necessary to conduct clinical evaluations of children for mental health or special education intervention, mental health diagnosis or special education classification. These professionals should also be well-versed in the tenets of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014).

If an SEL assessment will be used to make high-stakes decisions about a school or program, it is highly recommended that evidence be carefully scrutinized with the assistance of someone with psychometric expertise and an understanding of the context in which the decisions will be made.

### **Structure of the guide.**

This guide is divided into four parts and all four parts list questions that should be considered when evaluating the measurement quality of the SEL assessment. A table for each of the following four parts can be found in the [Appendices](#).

**Part 1: Does the assessment effectively measure the SEL competencies of interest?**

**Part 2: Does the SEL assessment provide credible evidence for your intended uses?**

**Part 3: Is the SEL assessment relevant for your students and your setting?**

**Part 4: Does the SEL assessment address issues relate to administration, scoring and assessment format?**

## Guide 3: Recommendations when Using Social and Emotional Learning (SEL) Assessments

Valid interpretation and use of social and emotional learning (SEL) assessment results does not end after identifying an assessment that aligns with your local purposes and ensuring the measurement quality. It also involves careful consideration of administration and scoring and communicating the results from an SEL assessment. This guide was developed using the [2014 Standards for Educational and Psychological Testing](#) published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) to identify considerations before and after administering an SEL assessment.

### Before Administering and Scoring the Assessments

Most SEL assessments have manuals that provide guidance on how to administer and score the assessment appropriately. Use this guidance to develop a formal plan for preparing personnel who will use the assessment. Outline the following considerations in that plan:

#### Identifying and sharing standardized procedures for administration and scoring.

Standardized administration procedures may involve instructions, time limits, and assessment conditions. Standardized scoring procedures may address how to aggregate item responses or apply a rubric. Follow any required qualifications or training recommended by the assessment developer for administration and scoring and check for correct and consistent administration and scoring across individuals, classrooms, or schools as well as documenting of deviations or disruptions.

If there is a need to ensure the assessment is being administered and scored correctly and consistently across individuals, classrooms, or schools, additional checks for quality control might be called for or documenting of deviations or disruptions.

#### Availability and use of accommodations

If students need alterations to administration and scoring procedures to receive full and fair access to the assessment (e.g. students with disabilities or from diverse cultural and linguistic backgrounds), identify valid accommodations documented by the assessment developer. If there is no documentation of a needed accommodation, consult state or school district policies for guidance.

Inform students as well as the school personnel administering the assessment about available accommodations and process for obtaining those accommodations.

### **Security and integrity of assessment materials and scores.**

Protect copyrighted materials by not allowing reproduction or re-creation of assessment materials in paper or electronic form without consent of the copyright holder. If results are viewed as consequential, have a plan for securing assessment materials as well as protecting the integrity of scores from fraud or deceit on the part of the respondent or assessment user.

### **Providing instructions, practice or other support to test takers.**

Inform respondents if the way in which they respond (e.g. guessing or the speed of their responses) could affect their scores. If unfamiliar equipment or software is used in administration (e.g., computer administered assessments), provide respondents practice opportunities with equipment or software unless the use of unfamiliar tools is part of what is being assessed (e.g., problem solving).

### **Collection of empirical evidence for alterations**

If altering the assessment format, mode of administration, instruction, or language of the assessment, the assessment developer should provide (or alternatively the test user should collect) empirical evidence that those alterations will not affect reliability/precision and validity of score interpretation. Published norms may not be valid under altered conditions if it is determined that changes to the assessment alter the meaning of scores.

## **Before Reporting and Interpreting Results from an SEL Assessment**

Test users should only report and interpret results as recommended and empirically supported by the assessment developer. To reach conclusions that validly inform decision-making, keep the following considerations in mind:

### **Levels at which results are reported.**

Report assessment scores only at the level intended and empirically supported by the assessment developer. For example, group vs. individual scores, overall scores vs. subscores, separate scores for subgroups. Never combine scores or separate scores unless recommended and empirically supported by the



assessment developer. If reporting subgroup results, individual users familiar with those subgroups should be involved in interpretation and use.

### **How results are reported.**

Ensure reporting of results protects copyright of the assessment materials and the privacy of assessment takers through security and confidentiality of individual scores. Consult the developer's cautions about the limitations of the scores, norms/comparison groups, and potential misinterpretation and misuse. Report amount of error expected for a score using standard error or confidence intervals to indicate scores are estimates that can vary from one occasion to the next.

Use simple language to describe what the assessment covers, what scores represent, the precision/reliability of the scores, and how to interpret and use scores. If reporting performance categories or labels, clearly and precisely describe the intended inference. Minimize potential biases for assessment takers due to demographics (e.g. cultural groups, age, social class, gender etc.)

Identify supplemental information (e.g., results from other assessments, academic/behavioral data) that would support appropriate interpretation and use, especially if reporting individual-level scores. Indicate how to weigh assessment scores in relation to supplemental information when making decisions.

If using an assessment regularly over time and/or used previously, verify that assessment interpretations remain relevant and appropriate when there are significant changes in the SEL curriculum/instruction, the population of assessment takers, modes of administration, or the purpose of conducting the assessment.

## **Before Communicating Assessment Results and Conclusions to Stakeholders**

A plan for communicating assessment results to stakeholders can assist in ensuring valid interpretation and use and minimize potential negative consequences. Considerations include the following.

### **Providing framing information.**

Provide assessment name, quotes of the purpose and intended interpretation and use, and cautions about interpretation and use from the assessment developer at the beginning of every discussion of assessment results. If sharing

assessment results publicly, accompany those results with enough information about the purpose of the assessment and how to appropriately interpret results to minimize the possibility of misinterpretations.

### **Anticipating misinterpretations and setting parameters for the discussion.**

Anticipate the possibility stakeholders might oversimplify their interpretations of results or misattribute reasons for results. Encourage sound conclusions and decision-making by thinking about these potential issues ahead of time. Before discussing assessment results, use recommendations from the assessment developer to define conversations about the results indicating what topics and conclusions are within bounds and outside of bounds (e.g. assigning meaning to results that were unintended or have no evidential basis).

## Appendices

Part 1: Does the assessment effectively measure the SEL competencies of interest?

Part 2: Does the SEL assessment provide credible evidence for your intended uses?

Part 3: Is the SEL assessment relevant for your students and your setting?

Part 4: Does the SEL assessment address issues relate to administration, scoring and assessment format?

# **Evaluating the Measurement Quality of Social and Emotional Learning (SEL) Assessments**

- 1. Does the assessment effectively measure the SEL competencies of interest?*

## Does the assessment effectively measure the SEL competencies of interest?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Clearly identify and define which SEL competencies the assessment measures.</p>	<p>Determine if SEL competencies of interest align with the SEL competencies measured by the assessment.</p>	<p>In order to determine whether an assessment will measure the SEL competencies of interest, those competencies must be stated in measureable terms that not only identify the competency of interest but also what a student will know, do, and/or understand as a result of achieving the SEL competency.</p> <p>SEL assessment should measure not only the competency of interest but also how students are expected to express that competency.</p> <ul style="list-style-type: none"> <li>• SEL competencies of interest could be more general (e.g. intrapersonal or interpersonal skills) or more specific (e.g. growth mindset, self-efficacy, collaborative problem solving).</li> <li>• Expression of SEL competencies might also differ such as demonstrating awareness (e.g. mindsets, knowledge, beliefs) or applying skills (e.g. learned abilities).</li> </ul> <p>For example, if students should demonstrate problem-solving skills, the assessment should measure how students use and apply those skills not whether they are aware of the importance of those skills.</p>	<p>If a measure addresses none of the specific or general SEL competencies of interest or very few, find another assessment.</p> <p>If the measure addresses some but not all SEL competencies of interest, look for a more comprehensive measure or a second measure to supplement information gathered.</p> <p>If an SEL assessment does not provide a clear description of SEL competencies measured, do a formal review of items/tasks to make your own determination or find another assessment that does measure the SEL competencies of interest.</p>

**Does the assessment effectively measure the SEL competencies of interest?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
Identify why intended respondents for the assessment are in the best position to assess students' SEL competencies.	Consider whether the respondent for the assessment is the best source for assessing the SEL competencies of students in the local population.	<p>If SEL competencies of interest involve attitudes, beliefs, or growth mindsets, respondents could be students reporting on their own SEL competencies.</p> <p>If the SEL competencies are behaviors, respondents should be individuals who know the students well enough to assess their SEL competencies.</p> <p>If the SEL competencies are knowledge or mental processes, responses should involve students demonstrating those SEL competencies through a direct assessment or performance task.</p>	If the intended respondents for the assessment are unfamiliar or unable to assess accurately SEL competencies in the local student population, find another assessment.

This space was intentionally left blank.

The table continues on the next page.

## Does the assessment effectively measure the SEL competencies of interest?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Use a representative panel of content experts to develop and/or review items/tasks and scoring protocols to ensure that the assessment addresses SEL competencies sufficiently and appropriately.</p>	<p>Conduct a local review of assessment items/tasks and scoring protocols to determine if those items sufficiently and appropriately address the competencies and outcomes for the local SEL program.</p>	<p>Clear and detailed specifications of the SEL competencies measured is important when developing not only tasks but also scoring protocols to ensure alignment between those defined specifications and the items/tasks and scoring protocols.</p> <p>Assessment developers can use expert review, an assessment blueprint, and/or mapping of items/tasks onto scores, to demonstrate that items/tasks represent a cross-section of competencies measured. For example,</p> <ul style="list-style-type: none"> <li>• Asking individuals with emotional regulation expertise to review items from an emotional regulation scale and indicate the extent to which each item aligns with the SEL competency and if the set of items overlook important aspects of the SEL competency.</li> <li>• Having a group of content experts review the number and content of items/tasks to determine if the assessment cover all measured SEL competencies sufficiently.</li> </ul> <p>As a general guideline,</p> <ul style="list-style-type: none"> <li>• Selected-response assessments should have at least three to five items for each competency measured.</li> <li>• Performance assessments typically involve a smaller number of tasks but that could hinder the generalizability of the scores if there is too broad of a set of SEL competencies measured.</li> </ul>	<p>If the developer does not document that the assessment sufficiently and appropriately addresses SEL competencies, conduct a local review with relevant expertise or find another assessment that provides this type of documentation and evidence.</p>

## Does the assessment effectively measure the SEL competencies of interest?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
Provide empirical evidence that items/tasks used to measure each competency are more highly related to each other than to items that measure other competencies (internal structure).	Determine if evidence supports that items/tasks used to measure SEL competencies are more highly related to each other than they are to items that measure other competencies.	<p>If an assessment claims to measure three competencies, there should be higher correlation among items/tasks that measure the same competency than among items/tasks that measure the other two competencies.</p> <p>Statistical analyses are used to support the assumption that unique rather than redundant information about each competency exists and these analyses typically require large sample sizes. For example,</p> <ul style="list-style-type: none"> <li>• For selected-responses assessments, confirmatory factor analysis can provide evidence that items load significantly on to factors that represent the different SEL competencies measured by the assessment.</li> <li>• For performance assessments, generalizability may be used to demonstrate that variability exists across different tasks.</li> </ul>	If evidence of internal structure does not support that items/tasks measuring a SEL competencies are unique rather than redundant of items/tasks measuring other SEL competencies, use caution when reporting, interpreting, and/or using scores for individual competencies.



# **Evaluating the Measurement Quality of Social and Emotional Learning (SEL) Assessments**

*2. Does the SEL assessment provide credible evidence for your intended uses?*

**Does the assessment provide credible evidence for intended uses?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Clearly state the intended interpretation and uses for the assessment score(s) and highlight evidence that justifies using the assessment for those interpretations and uses.</p>	<p>Ensure that the assessment developer's stated interpretations and uses align with local plans for using assessment results and determine if evidence supports those interpretations and uses.</p>	<p>Measures might be developed for screening, formative, interim, and/or summative purposes, and this intent should be specified by the assessment developer and align with local plans for using the data. For example,</p> <ul style="list-style-type: none"> <li>• If teachers will use the information to guide instruction, then use a formative assessment measure that provides classroom-level data to guide those instructional decisions.</li> <li>• If a school plans to use an assessment in an improvement process, then use an interim or summative measure that provides school-level data to assess progress and determine how to move forward.</li> </ul>	<p>If the assessment developers' intended interpretations and uses for an SEL assessment do not align with local plans or are unsupported, find another assessment that does align with plans for use.</p>

This space was intentionally left blank.

The table continues on the next page.

## Does the assessment provide credible evidence for intended uses?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Identify score(s) provided (e.g. overall score, subscores, performance levels) and items/tasks used to generate each score.</p> <p>Clearly state recommendations and limitations for reporting and interpreting those scores.</p>	<p>Determine if scores provided will guide intended uses or assist in reaching conclusions about students' achievement of SEL competencies.</p> <p>Ensure that local plans for reporting and interpreting assessment results follow developer's recommendations and limitations.</p> <p>Be alert to possible misinterpretation of scores and take steps to minimize inappropriate interpretation and use.</p>	<p>Do not interpret assessment results for purposes unless recommended by the developer with the support of evidence. Examples include:</p> <ul style="list-style-type: none"> <li>• Most SEL competency assessments are appropriate for assessing students' strengths and <u>do not</u> have enough evidence to support using the assessment for screening or diagnosing mental health issues.</li> <li>• If the assessment reports multiple scores, do not aggregate those into a single score unless the developer provides evidence that doing so is appropriate.</li> <li>• If the assessment reports a single composite score, do not disaggregate the score unless the developer provides evidence that doing so is appropriate.</li> <li>• If the assessment will guide instruction or practice, reported scores should provide enough specificity to inform these intended uses such as by providing subscores on specific domains or competencies.</li> <li>• If an assessment will determine whether SEL has occurred, an SEL program is effective, or whether SEL learning goals are met, reported scores could be more general.</li> </ul> <p>Holistic and analytical scoring are typical for many performance assessments.</p> <ul style="list-style-type: none"> <li>• For holistic scoring, results are a single, holistic judgement about a students' SEL.</li> <li>• In analytical scoring, decisions result in judgements about one or more SEL competencies. Analytical scoring potentially can provide more information about strengths and weaknesses but requires evidence that those scores are able to differentiate between different SEL competencies.</li> </ul>	<p>If scores provided by the assessment will not guide intended uses or inform conclusions at the local level, find another assessment.</p> <p>Do not attempt to combine or calculate scores from an assessment without proper psychometric evidence.</p> <p>If assessment developer's recommendations and cautions for reporting or interpreting SEL assessment results do not align with local plans for reporting and interpretation, find another assessment that aligns with local plans.</p>

**Does the assessment provide credible evidence for intended uses?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
Cite theory, research, or empirical evidence that students/observers/interviewers interpret and respond to items/tasks as intended.	Review rationale or evidence provided by the assessment developer that respondents respond as intended to determine if it supports the use of the assessment with the local population and setting.	<p>Assessments should find a way to document that respondents are answering items/tasks using the processes and behaviors the developer intended. For example,</p> <ul style="list-style-type: none"> <li>• Interviewing respondents about their response choices as they complete items.</li> <li>• Collecting feedback from raters about the factors they considered when assigning their ratings.</li> </ul>	If there is insufficient rationale or evidence that respondents are interpreting and responding as intended, use other evidence of SEL competencies to confirm interpretations.

This space was intentionally left blank.

The table continues on the next page.

**Does the assessment provide credible evidence for intended uses?**

*If the assessment will be used to determine students' strengths and needs...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Provide empirical evidence of consistency of item results (internal reliability) for all assessment scores reported.</p>	<p>Determine if assessment scores have an acceptable reliability coefficient (.80 or above for coefficient alpha).</p>	<p>Consider reliability evidence for each score to be reported understanding that aggregating scores at a class, group, grade, or school level will be more reliable than scores for individual students.</p> <p>If validity evidence appears to support assessment at the individual student level, a measure of internal consistency will indicate the extent to which a respondent responds similarly across items.</p> <p>Internal reliability typically takes the form of a coefficient alpha.</p> <ul style="list-style-type: none"> <li>• Coefficient alpha ranges between 0 and 1 with a value closer to 1 indicating better consistency (reliability).</li> <li>• The stakes of an intended use is a basis for determining the degree of reliability required, with higher reliability needed when stakes are higher.</li> <li>• A minimum threshold for reliability is .80. Reliability slightly below .80 is undesirable but may not be problematic. Reliability significantly below .80 is problematic for interpretation and use.</li> </ul> <p><b>NOTE:</b> Sufficient reliability evidence is not enough to support the use of scores to make consequential decisions about individual students, such as for diagnosis or program placement.</p>	<p>If the internal reliability of any score reported is below .80, even slightly use caution when interpreting and using those scores for decisions about individual students.</p> <p>If the internal reliability of any score is not reported or considerably below .80, do not report, interpret, and/or use any scores/subscores that do not meet this minimum or find an assessment where all scores reported are sufficiently reliable.</p>

**Does the assessment provide credible evidence for intended uses?**

*If the assessment will be used to determine students' strengths and needs (continued)...*

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
Provide a standard error of measurement and recommended confidence intervals/bands for all reported assessment scores.	When reporting and interpreting scores, include some reference to the true range of those scores based on standard error of measurement and confidence intervals or bands.	<p>If an assessment provides evidence that supports reporting individual scores, also report confidence intervals to capture the true potential range of the students' performance.</p> <p>Confidence intervals are particularly important when comparing two different scores. For example,</p> <ul style="list-style-type: none"> <li>• Comparing an individual student's score against a criterion score such as proficiency level or norms.</li> <li>• Comparing changes in an individual's score over time.</li> <li>• Comparing the scores of two different individuals.</li> </ul>	<p>If standard error of measurement and/or confidence intervals or bands are not available,</p> <ul style="list-style-type: none"> <li>• contact the developer for this information,</li> <li>• use caution when determining students' strengths and needs, and/or</li> <li>• double check with other information about students' SEL competencies to see if the two sources agree.</li> </ul>
<p><i>See also expectations for "Is the assessment relevant for the students and the setting?"</i></p>			

**Does the assessment provide credible evidence for intended uses?**

*If the assessment will be used to compare scores over time...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
Provide empirical evidence that scores are sensitive to changes in SEL over time.	Determine if evidence is applicable to the local setting and program and provides supportive evidence that the assessment will capture changes in SEL that occur over time.	Typically, cross sectional and longitudinal studies provide evidence that the scores of an assessment given at two different points in time would reflect a change in SEL if such a change did occur. <ul style="list-style-type: none"> <li>• For example, comparing SEL skills at the beginning and end of the school year after students completed the SEL program.</li> </ul>	If sensitivity to change over time is unsupported, do not use the assessment to determine if change over time has occurred.

*If the assessment will be used to evaluate an SEL Program...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
Provide evidence that assessment score(s) demonstrate change after implementing an SEL program that has been shown to be effective at improving the competencies measured by the assessment.	Determine if evidence provides information that is applicable to the local setting and program.	Evidence of how sensitive an assessment is to change could involve a field testing study. <ul style="list-style-type: none"> <li>• For example, students who received instruction or maybe even higher quality instruction would score significantly higher on the assessment than students who did not.</li> </ul>	If there is insufficient evidence that assessment scores can demonstrate change, be cautious about using scores to evaluate the effectiveness of the SEL program and/or instruction.

**Does the assessment provide credible evidence for intended uses?**

*If the assessment will be used to improve school/program quality...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
Provide evidence that assessment score(s) are moderately related to desirable educational outcomes (e.g. graduation, absentee rates, etc.)	Determine if evidence provided is applicable to the local quality improvement goals or outcomes.	Longitudinal, quasi-experimental, or experimental research studies can be used to determine if there is a significant correlation between relevant indicators of quality and the assessment score.	If there is insufficient evidence that score(s) are highly related to quality outcomes of local interest, do not use scores to make decisions about improving school/program quality.

This space was intentionally left blank.

The table continues on the next page.



## Does the assessment provide credible evidence for intended uses?

*If the assessment will be used to report separate results for different groups of students...*

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Provide rationale or evidence that students from different groups conceptualize, define, and experience the SEL competencies assessed by the assessment.</p>	<p>Review rationale or evidence provided to determine applicability to the local setting, SEL program, and demographics of the local student population.</p>	<p>If using the results of an SEL assessment to report separate results for different groups of students, it is important to ensure that relevant groups of student experience the assessed SEL competencies similarly.</p> <ul style="list-style-type: none"> <li>• For example, if reporting results separately for different racial/ethnic groups then the competencies measured should be culturally relevant for students in the local student population.</li> </ul> <p>If group difference are reported, do so cautiously and only after thorough review.</p>	<p>If there is insufficient rationale or evidence different groups of students conceptualize define, and experience SEL competencies similarly,</p> <ul style="list-style-type: none"> <li>• ask individuals from representative groups to review the relevance of SEL competencies assessed, or</li> <li>• do not report and compare results for different groups of students.</li> </ul>
<p>Provide evidence that assessment score(s) are equally valid, reliable, and fair for different groups of students.</p> <p>If not, clearly caution against the reporting of assessment scores for groups of students separately.</p>	<p>Determine if evidence provided is applicable to the local setting, SEL program, and demographics of the local student population and supports reporting scores separately for different groups of students.</p>	<p>Because of potential issues with relevance of SEL assessments for different groups of students (e.g. cultural, gender, age), if schools have an interest in comparing or reporting separately the results for different groups of students the school should:</p> <ul style="list-style-type: none"> <li>• Justify the use of those results for solving a specific problem of practice rather than just using it to report how different groups perform.</li> <li>• Ensure validity, reliability, and fairness study samples include students from different groups that will be compared or results reported separately. Preferably, require validity, reliability, and fairness study results are report separately for different groups of students.</li> </ul>	<p>If there is insufficient empirical evidence that score(s) are valid, reliable, and fair for different groups of students or the assessment developer cautions against it, do not report and interpret scores for groups of students separately.</p>

# **Evaluating the Measurement Quality of Social and Emotional Learning (SEL) Assessments**

*3. Is the SEL assessment relevant for your students and your setting?*

## Is the assessment relevant for the students and the setting?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Identify the intended population for the assessment and clearly articulate if there are any inclusion or exclusion criteria.</p>	<p>Select an assessment that is intended for the key demographics (e.g. age/grade) of the local student population to be assessed.</p>	<p>Use assessments only with individuals who are demographically representative of the intended population. For example,</p> <ul style="list-style-type: none"> <li>• Do not use an assessment developed for Grades 9 and up if the intended population of the assessment is middle or elementary school students</li> <li>• Do not use an assessment with English Learners (ELs) if a developer indicates that the assessment is not appropriate for those students.</li> </ul>	<p>If the intended population for the assessment does not align with the key demographics of the local student population to be assessed, look for another assessment.</p>
<p>Provide a rationale and evidence that what and how SEL competencies are measured is developmentally appropriate for the grades/ages of students in the intended population.</p>	<p>Review the rationale and evidence to determine if the assessment is developmentally appropriate for the grade/ages of students in the local population.</p>	<p>Developmental appropriateness is particularly important if an assessment will be used to track SEL competency development over ages or grades.</p> <p>Student development of SEL competencies can differ not only because:</p> <ul style="list-style-type: none"> <li>• Different SEL competencies become important at different developmental stages.</li> <li>• Ways in which those SEL competencies are demonstrated or displayed changes over time.</li> </ul> <p>An assessment developer should address these developmental considerations when developing and validating the assessment.</p>	<p>If there is an insufficient rationale or evidence that an assessment is developmentally appropriate for the grades/ages of students in the local population, use for the grades/ages for which it would be appropriate or find another assessment.</p>

## Is the assessment relevant for the students and the setting?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
Indicate the reading level and linguistic competency needed by respondents.	Determine if respondents will have appropriate levels of reading and linguistic competence.	<p>The reading level and linguistic complexity of an assessment is not only important to consider in terms of students but for other respondents as well.</p> <ul style="list-style-type: none"> <li>• For example, if a parent report would require a sixth grade reading level and English proficiency, ensure that most if not all parents will meet those requirements; if not, determine how to accommodate the participation of parents who do not meet those requirements.</li> </ul>	<p>If an assessment developer does not specify reading or linguistic competency needed by respondents,</p> <ul style="list-style-type: none"> <li>• ask the assessment developer for more information, or</li> <li>• have a reading specialist/ELL coordinator review the assessment to determine if it is appropriate for the local population.</li> </ul>
<p>Specify the availability of language and ability accommodations are available.</p> <p>For available accommodations, provide guidance on when to use the accommodation and how to administer and score it.</p>	Determine if accommodations for students or respondents in the local population are available.	<p>If the setting has a linguistically diverse student population or a sizable number of students with identified disabilities, the availability of accommodations would allow these students to participate.</p> <p>Seeking out the availability of multi-language versions or forms for students with disabilities would be another option.</p>	<p>If needed accommodations are not available, ask assessment developer for more information.</p> <p>If adequate accommodations do not exist,</p> <ul style="list-style-type: none"> <li>• do not use the assessment for relevant students, or</li> <li>• seek out experts who can assist in identifying accommodations that would remove barriers for these students but not change the SEL competencies measured.</li> </ul>

**Is the assessment relevant for the students and the setting?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Use a culturally representative panel to review SEL competencies measured by the assessment to determine if how those SEL competencies are measured are relevant for different cultures.</p>	<p>Review the demographics and findings of the panel to insure individuals from cultural groups represented in the local setting and student population are included and the assessment will fairly assess SEL competencies for those cultural groups.</p>	<p>How an SEL assessment defines and measures competencies may not be relevant to respondents from different cultural groups because the value of those SEL competencies or how they are represented may vary.</p> <p>Cultural differences are an important consideration when developing SEL programs and assessments along with systematic review and/or empirical studies to ensure they are not culturally biased. For example,</p> <ul style="list-style-type: none"> <li>• Review panels should include members of each relevant cultural group or people either who work with or are familiar with those groups.</li> <li>• Comments from such individuals should be considered seriously.</li> </ul>	<p>If the SEL competencies addressed by the assessment have not been reviewed and approved by a culturally representative panel,</p> <ul style="list-style-type: none"> <li>• ask a panel that represents cultural groups in local student population about the relevancy of the SEL competencies, or</li> <li>• do not use the assessment for making decisions about unrepresented cultural groups.</li> </ul>

This space was intentionally left blank.

The table continues on the next page.

**Is the assessment relevant for the students and the setting?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Have a diverse panel familiar with the needs of different students review the content and format of the assessment for bias, sensitivity, and accessibility.</p> <p>Document whether that panel found with a high level of agreement that assessment is unbiased, sensitive, and accessible.</p> <p>If the panel identifies items or format as biased, insensitive, or inaccessible, describe how those issues were addressed.</p>	<p>Review the demographics of the panel and the findings of the panel to determine if the review is applicable to local setting and student population and if any bias issues were raised that might be a concern for the local student population.</p>	<p>Individuals of different backgrounds or individuals who are aware of capability differences among students should be involved in the development and review of SEL assessments. This includes:</p> <ul style="list-style-type: none"> <li>• Review panels representing different racial/ethnic groups, ages, gender, individuals with disabilities, etc.</li> <li>• Reviewing items for topic and wording that could be unfair or ratings of students by individuals who might have an unconscious bias.</li> </ul>	<p>If the developer has not used a panel that is representative of the local student population to review for bias, sensitivity, and accessibility, ask a local group that is familiar with the needs of different students in the local population to review the assessment and its items.</p> <p>If there is insufficient documentation that an assessment will be fair for specific demographic groups, do not use the assessment for those demographic groups.</p>

This space was intentionally left blank.

The table continues on the next page.

## Is the assessment relevant for the students and the setting?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Provide empirical evidence that responses to items/tasks and reported scores are not significantly different for students with similar levels of SEL competency from different demographic groups (e.g. race/ethnicity, language, and gender).</p> <p>If statistical differences exist, indicate actions taken to understand those potential differences better.</p>	<p>Review the evidence provided to determine if the assessment addresses key student groups in the local population and if it raises fairness concerns for individuals from those groups.</p> <p>Demographics to consider include gender, race/ethnicity, socioeconomic status, and language background.</p>	<p>Assessments can consist of items/tasks that do not function the same way for different group of students or differences between relevant subgroups on reported scores.</p> <p>At the item, measurement invariance studies (e.g. differential item functioning or multigroup confirmatory factor analyses) gather evidence that assessment items performs the same way for different groups of students.</p> <ul style="list-style-type: none"> <li>• If studies find a lack of measurement equivalence, a follow-up study determining whether differences are potentially due to bias should occur before a user can credibly use the assessment for measuring the SEL competencies of those diverse groups.</li> <li>• Those conducting the analyses must also be aware of the assumptions of the procedures and number of individuals needed to conduct those analyses to avoid misinterpretation of results.</li> </ul> <p>At the score level, differential prediction is a common method used to determine through regression analysis whether there are differences between relevant subgroups on reported scores.</p>	<p>If there is insufficient evidence that students with similar levels of SEL competency from demographic groups respond at the item/task or score level similarly,</p> <ul style="list-style-type: none"> <li>• do not report and compare the scores from subgroups, or</li> <li>• find another valid assessment for student populations that are demographically very diverse.</li> </ul> <p>If there is evidence of lack of equivalence at the item or score level that was not addressed by the assessment developer, do not interpret and use assessment results for those subgroups especially if they are a key group in the student population.</p>

# Evaluating the Measurement Quality of Social and Emotional Learning (SEL) Assessments

- 4. Does the SEL assessment address issues related to administration, scoring, and the assessment format?*



**Does the assessment address issues related to administration, scoring and the assessment format?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Provide detailed and clear instructions if test users will administer and score the assessment.</p> <p>If applicable, indicate if there are specific qualifications or training experiences needed to administer and score the assessment.</p>	<p>Ensure that all individuals administering and scoring the assessment receive instructions provided by the assessment developer.</p> <p>If applicable, ensure qualified or trained individuals are available to administer and score.</p>	<p>Logistics and required training time should be considered when making decisions to use a particular assessment. Training of the following individuals might be necessary:</p> <ul style="list-style-type: none"> <li>• Individuals administering assessments, completing rating scales, or conducting observation may need training on how to complete the assessments.</li> <li>• Individuals compiling and reporting data may need training on developer recommendations for reporting, interpretation, and use.</li> <li>• Individuals who will use and communicate findings might also need training such as how to communicate findings to students and families.</li> </ul> <p>Some assessments require that those administering and/or scoring an assessment have certain qualifications such as a degree, graduate coursework, or specific formal training.</p> <p>Even if an assessment does not have requirements for administration and scoring, consider guidance that encourages standardized administration and scoring for comparable scores.</p>	<p>If requirements for administration and scoring are unaddressed in the assessment documentation, ask the assessment developer for more information.</p> <p>Do not use the assessment if qualified individuals are not available or training of individuals to administer and score the assessment would not be possible.</p>

**Does the assessment address issues related to administration, scoring and the assessment format?**

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>If the test developer administers or scores the assessment, describe the process for conducting the assessment and/or the procedure used for generating scores.</p>	<p>Ensure that the basis for administering items and/or generating scores aligns with definitions for SEL competencies and supports local plans for interpretation and use.</p>	<p>Some test developers will use automated means for administering or scoring assessments that often involve algorithms.</p> <p>Algorithms for scoring assessments or selecting items can be very technical, but developers should be able to explain conceptually how the algorithm works.</p> <p>This conceptual explanation will help indicate whether the assessment's administration and scoring procedures are appropriate for the local setting and SEL program.</p>	<p>If there is insufficient information about how the assessment is administered and scored, ask the developer for more information.</p> <p>If administration and scoring procedures are not appropriate for the local setting, student population, or SEL program, find another assessment.</p>
<p>Indicate if specific technological devices and software to administer and/or score the assessment are required or recommended.</p>	<p>Ensure that the all settings (e.g. schools) administering the assessments have access to required or recommended technological devices and software.</p>	<p>If administering an assessment via a technological device, there likely are requirements for the devices and type of software available on those devices.</p> <p>Differences in mode (e.g. paper and pencil vs. computer-delivered), device (e.g. desktop computer vs. tablet), or operating system (e.g. Windows vs. Macintosh) could differentially affect how assessments are completed by respondents and compromise score comparability.</p>	<p>If the required devices or software are not available, find another assessment.</p> <p>If not all settings administering the assessment have access to recommended technological devices and software,</p> <ul style="list-style-type: none"> <li>• find another assessment,</li> <li>• do not use the assessment in those settings, or</li> <li>• request evidence from the assessment developer that differences in devices or software used to administer or score the assessments will not affect score comparability.</li> </ul>

## Does the assessment address issues related to administration, scoring and the assessment format?

### *If assessment scores are determined using norms...*

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Report norms should be:</p> <ul style="list-style-type: none"> <li>• based on a recent, representative sample of sufficient size,</li> <li>• document the demographics of the students included in the sample (e.g. gender, age/grade, race/ethnicity, SES, geographic location), and</li> <li>• describe the setting in which the norm data were gathered.</li> </ul>	<p>Ensure the norm study and sample is:</p> <ul style="list-style-type: none"> <li>• current (gathered in last 5-7 years),</li> <li>• of sufficient size (500 or more total and 100 or more per grade/age group),</li> <li>• gathered from a setting similar to the local setting, and</li> <li>• collected from a student sample that includes representation of the local student population (e.g. gender, race/ethnicity, SES, geographic location).</li> </ul>	<p>Norm samples should include and document:</p> <ul style="list-style-type: none"> <li>• A proportional representation of students from different demographic groups (note number of English Learners in the sample).</li> <li>• The relevant setting in which a norm sample was administered the assessment.</li> </ul> <p>For example, norms developed using a predominately students from urban high school would not be relevant for rural middle school students.</p>	<p>If the norm sample is not current, is not of sufficient size, or does not represent students from different demographic groups relevant to the local population,</p> <ul style="list-style-type: none"> <li>• ask the developer about the availability of updated and relevant norm information,</li> <li>• do not use the norm-referenced scores for reporting or decision-making, or</li> <li>• find another assessment with applicable norms.</li> </ul>

### *If there are multiple forms (different versions) for an assessment (e.g. Forms A & B)...*

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<p>Provide evidence of score consistency across the different forms.</p>	<p>Determine if the evidence supports that scores from different forms of the assessment are comparable.</p>	<p>Equating is a commonly used technical process that establishes scores are interchangeable across different versions of a test.</p> <p>Equating samples need to be large and representative of the population under consideration for assessment.</p>	<p>Only use one form of the assessment if there is insufficient evidence that scores from multiple forms would provide consistent results across students.</p>

**Does the assessment address issues related to administration, scoring and the assessment format?**

*If the assessment is a completed by a student...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Indicate how development or administration of the SEL assessment addresses common issues such as memory bias, social desirability bias, or reference bias.</p>	<p>Determine if the developer has provided convincing evidence or rationale that the SEL assessment is not susceptible to these biases.</p>	<p>Memory, social desirability, and reference biases are common issues to address in the development or administration of assessments where the student is the respondent.</p> <ul style="list-style-type: none"> <li>• Memory bias occurs if respondents are not aware or accurate in the assessment of their SEL behaviors or actions.</li> <li>• Social desirability bias involves the respondent providing an answer considered attractive instead of what is true for him/her.</li> <li>• Reference bias are responses affected by whom respondent compares his/her SEL competence. Such as, if an assessment has consequential decisions for students, they also may not be inclined to answer accurately.</li> </ul>	<p>If there is insufficient evidence or rationale for how potential biases were addressed or mitigated in development or administration,</p> <ul style="list-style-type: none"> <li>• ask the assessment developer for more information, or</li> <li>• ask a small group of potential respondents or individuals familiar with respondents to review items and determine if these biases could be problematic.</li> </ul>

This space was intentionally left blank.

The table continues on the next page.

**Does the assessment address issues related to administration, scoring and the assessment format?**

*If the assessment is a rating or observation scale completed by someone other than the student...*

<b>Assessment developer should...</b>	<b>Test user should...</b>	<b>Explanation</b>	<b>What to do if an assessment does not meet this criterion?</b>
<p>Provide evidence that the administration and scoring protocol will lead to consistent decisions across different raters/observers (interrater reliability) and avoid or mitigate potential biased ratings</p>	<p>Use recommended training and protocols to avoid or mitigate biases.</p> <p>Determine if interrater reliability is acceptable (Kappa or Intraclass Correlation Coefficient (ICC) statistic of .70 or higher).</p>	<p>These types of assessment should provide evidence of interrater reliability because some teachers might rate differently than other teachers across items/tasks or students. Common rating issues include:</p> <ul style="list-style-type: none"> <li>• Inclination to rate students they "like" more positively than other students (halo effect).</li> <li>• Use more leniency or severity in ratings.</li> <li>• Misinterpret/misattribute sources of behavior.</li> <li>• Rating accuracy affected if respondents have a personal or professional stake in the results of the assessment (e.g. evaluate teacher performance).</li> </ul> <p>Such disparities would affect the consistency across raters. Therefore, these types of assessments should provide instructions on how to help raters/observers overcome these response biases.</p> <ul style="list-style-type: none"> <li>• For example, training observers on actual students, vignettes or videos with discussion of differences in ratings may be quite productive for calibrating ratings.</li> </ul>	<p>If there is insufficient information about how to avoid or mitigate rater response bias,</p> <ul style="list-style-type: none"> <li>• ask assessment developer for more information, or</li> <li>• ask a small group of potential respondents to review items and determine if these biases could be an issue for them or others.</li> </ul> <p>If there is insufficient evidence of interrater reliability or interrater reliability is considerably below .70,</p> <ul style="list-style-type: none"> <li>• ask the assessment developer for more information,</li> <li>• consider more training for raters/observers, or</li> <li>• find another assessment.</li> </ul>