

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

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<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"> • Individuals administering assessments, completing rating scales, or conducting observation may need training on how to complete the assessments. • Individuals compiling and reporting data may need training on developer recommendations for reporting, interpretation, and use. • Individuals who will use and communicate findings might also need training such as how to communicate findings to students and families. <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?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<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"> • find another assessment, • do not use the assessment in those settings, or • request evidence from the assessment developer that differences in devices or software used to administer or score the assessments will not affect score comparability.

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

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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<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"> • Memory bias occurs if respondents are not aware or accurate in the assessment of their SEL behaviors or actions. • Social desirability bias involves the respondent providing an answer considered attractive instead of what is true for him/her. • 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. 	<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"> • ask the assessment developer for more information, or • ask a small group of potential respondents or individuals familiar with respondents to review items and determine if these biases could be problematic.

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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<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"> • Inclination to rate students they "like" more positively than other students (halo effect). • Use more leniency or severity in ratings. • Misinterpret/misattribute sources of behavior. • Rating accuracy affected if respondents have a personal or professional stake in the results of the assessment (e.g. evaluate teacher performance). <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"> • For example, training observers on actual students, vignettes or videos with discussion of differences in ratings may be quite productive for calibrating ratings. 	<p>If there is insufficient information about how to avoid or mitigate rater response bias,</p> <ul style="list-style-type: none"> • ask assessment developer for more information, or • ask a small group of potential respondents to review items and determine if these biases could be an issue for them or others. <p>If there is insufficient evidence of interrater reliability or interrater reliability is considerably below .70,</p> <ul style="list-style-type: none"> • ask the assessment developer for more information, • consider more training for raters/observers, or • find another assessment.