

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?
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.	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>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"> • 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. • 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. 	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.

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"> • 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. • If the assessment reports multiple scores, do not aggregate those into a single score unless the developer provides evidence that doing so is appropriate. • If the assessment reports a single composite score, do not disaggregate the score unless the developer provides evidence that doing so is appropriate. • 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. • 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. <p>Holistic and analytical scoring are typical for many performance assessments.</p> <ul style="list-style-type: none"> • For holistic scoring, results are a single, holistic judgement about a students' SEL. • 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. 	<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?

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
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"> • Interviewing respondents about their response choices as they complete items. • Collecting feedback from raters about the factors they considered when assigning their ratings. 	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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
<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"> • Coefficient alpha ranges between 0 and 1 with a value closer to 1 indicating better consistency (reliability). • The stakes of an intended use is a basis for determining the degree of reliability required, with higher reliability needed when stakes are higher. • 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. <p>NOTE: 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"> • Comparing an individual student's score against a criterion score such as proficiency level or norms. • Comparing changes in an individual's score over time. • Comparing the scores of two different individuals. 	<p>If standard error of measurement and/or confidence intervals or bands are not available,</p> <ul style="list-style-type: none"> • contact the developer for this information, • use caution when determining students' strengths and needs, and/or • double check with other information about students' SEL competencies to see if the two sources agree.
<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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
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.	<p>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.</p> <ul style="list-style-type: none"> • For example, comparing SEL skills at the beginning and end of the school year after students completed the SEL program. 	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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
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.	<p>Evidence of how sensitive an assessment is to change could involve a field testing study.</p> <ul style="list-style-type: none"> • For example, students who received instruction or maybe even higher quality instruction would score significantly higher on the assessment than students who did not. 	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...

Assessment developer should...	Test user should...	Explanation	What to do if an assessment does not meet this criterion?
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"> • 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. <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"> • ask individuals from representative groups to review the relevance of SEL competencies assessed, or • do not report and compare results for different groups of students.
<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"> • Justify the use of those results for solving a specific problem of practice rather than just using it to report how different groups perform. • 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. 	<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>