

STEP Literacy Assessment

Criterion	Specific Indicators	Rating	Feedback from Reviewers	Tally of rating
Validity, Reliability and Consistency in Scoring				
Evidence of test reliability and consistency in scoring	<p>Results of reliability studies are reported for each grade assessment</p> <p>Evidence includes: The studies are appropriate given the purpose of the measure. For each grade-level, studies provide evidence of:</p> <ul style="list-style-type: none"> • Split-half reliability • Coefficient alpha • Test-retest reliability • Classification consistency 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. Correlations demonstrate ranges of .7 or higher. (2)</p>	<p>Data is limited with much narrative about the philosophy around the test, very little data, mainly on p. 61. Limited table on p. 27.</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>
	<p>Standard error of measurement or standard estimate of error is reported</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • SEM estimates are reported for score 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not</p>	<p>I found no evidence.</p>	<p>Does not meet: I</p> <p>Partially meets: I</p>

	<p>ranges and cut-scores.</p> <ul style="list-style-type: none"> SEM estimates are reported for score ranges and cut-scores for each assessment (grade-level, form, subtest). 	<p>demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS -- Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Meets or exceeds:</p>
	<p>Inter-rater reliability studies have been conducted. Study sample used to establish inter-rater reliability represents test administrators.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> Inter-rater reliability studies have been conducted for each grade level and are based on a representative sample of educators who will administer and score the assessment. Inter-rater reliability coefficients exceed .7. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS --most information for the criterion is provided. Information and data provided suggests acceptable or strong</p>	<p>The publisher request that the inter-rater reliability was not given in the best of circumstances. No data was provided. No grade level or student demographic information was provided.</p> <p>1 year, six schools, same city, 421 sample size</p>	<p>Does not meet: I</p> <p>Partially meets: I</p> <p>Meets or exceeds:</p>

		evidence. (2)		
	<p>Studies have been conducted to establish reliability with all subcategories of students who will take the assessment.</p> <p>Evidence Includes: Studies that demonstrate reliability has been established from scoring samples of students that include: Non-ELLs with and without reading deficiencies and ELLs with and without reading deficiencies.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>STEP provides limited information for the coefficients' ranging from .9 to .6. No evidence provided as to the population and demographics of tested population. 6</p> <p>4 cohorts from one school were mentioned</p>	<p>Does not meet: I</p> <p>Partially meets: I</p> <p>Meets or exceeds:</p>
Alternative forms available for multiple assessments with demonstrated equivalence or comparability	<p>If alternative forms are provided, all forms have demonstrated evidence of equivalence or comparability such as test-retest, parallel form and internal consistency.</p> <ul style="list-style-type: none"> Technical reviews indicate all forms for each grade level have demonstrated evidence of comparability and content specifications. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or</p>		<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

	<p>Evidence includes:</p> <ul style="list-style-type: none"> • Sufficient forms are provided to allow for progress monitoring between interim assessments. • Split-half reliability. • Coefficient alpha reliability. 	<p>data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence correlations demonstrate ranges of .7 or higher. (2)</p>		
Content and Construct Validity				
Evidence of content and construct validity	<p>Evidence reported to demonstrate the assessment helps correctly identify students with “<i>significant reading deficiencies</i>” so that successful remediation and intervention can be provided; studies have been conducted with similar assessments to show that the assessment measures reading ability, not other irrelevant criteria.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • A clear description is provided that demonstrates the purpose of the assessment is to screen students for reading concerns. • Content specifications for each grade-level, including a complete description of the test 	<p>Rating</p> <p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided</p>	<p>Could not locate evidence for a significant reading deficiency. Grade level correlations are made</p> <p>specifications by grade level given, but lack of research/evidence to support the criteria</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

	content, purpose(s), and intended use(s), and assessment blueprint as appropriate, is provided.	suggests acceptable or strong evidence. (2)		
	<p>Reading levels are reported for passages and how levels were established. Reading levels of assessment passages have been field-tested or have other evidence.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> Field testing populations should be clear and should mirror the school/district demographics. Statistics used to establish the reading levels are reported with both ELL and Non-ELL populations. Findings from a content review by field experts, including teachers in tested grade levels. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>No content review by field experts, field tested populations are not significant Lexile comparison was noted</p>	<p>Does not meet: I</p> <p>Partially meets: I</p> <p>Meets or exceeds:</p>
	If appropriate, findings from alignment studies to demonstrate alignment with Colorado Academic Standards for Language Arts and resolution for any resulting concerns.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the</p>	<p>No evidence or mention of CAS.</p> <p>No findings from studies noted in regards to alignment to standards</p>	<p>Does not meet: II</p> <p>Partially meets:</p> <p>Meets or exceeds:</p>

		<p>criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		
	<p>There are studies of construct validity, such as convergent and discriminant analysis, demonstrating correlations of .7 or above.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>Some data available; incomplete</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>
Evidence of	Evidence reported to	DOES NOT		Does not

<p>criterion/predictive validity accurately identifying students with “<i>significant reading deficiency</i>”</p>	<p>demonstrate that the assessment has established criterion and/or predictive validity to correctly identify students with and without a “<i>significant reading deficiency.</i>”</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • A clear definition of the criterion or measure that were used to establish concurrent validity. • Studies with similar assessments that demonstrate the assessment measures reading ability, not other irrelevant criteria. Predictive validity correlations above .7. 	<p>MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>Provided evidence of validity correlation from 2nd grade to the end of year testing at 3rd grade. This coefficient score(s) were ITBS .58 and ISAT .68. Below the cut score of .7.</p>	<p>meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>
<p>Determination of cut-scores based upon well-designed pilot study</p>	<p>The assessment has established cut-scores for decision making about students’ “<i>significant reading deficiency</i>” using adequate demographics representing (i.e., 10% ELL and 25% F/R lunch), appropriate criterion assessment, adequate sample size, and appropriate statistics.</p> <p>Evidence indicates:</p> <ul style="list-style-type: none"> • Includes a description of the process used to establish the cut points. • A full description of the 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p>	<p>Unable to pinpoint hard evidence to back this to a 2.</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

	<p>norming sample.</p> <ul style="list-style-type: none"> The norming sample is a large representative national sample of students at the same grade level and is representative of the testing population according to gender, ELL status, special needs status and F/R lunch status. 	<p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and 2data provided suggests acceptable or strong evidence. (2)</p>		
	<p>Studies of classification accuracy analysis provide evidence that the measure appropriately identifies students as indicated in the description of purpose of the assessment, demonstrating values that exceed .8 or higher.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Does not meet: Partially meets: II Meets or exceeds:</p>
	<p>Acceptable, recognized procedures are followed for setting cut-scores.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or</p>	<p>Some rationale.</p>	<p>Does not meet: Partially meets: II</p>

		<p>information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Meets or exceeds:</p>
	<p>SEM estimates are reported for cut-scores with guidance for score interpretation.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence.(1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and</p>		<p>Does not meet: Partially meets: II Meets or exceeds:</p>

		data provided suggests acceptable or strong evidence. (2)		
Universal Design	<p>Evidence reported to demonstrate that the assessment has cultural validity, that fairness and bias issues have been addressed; the assessment is accessible to all learners, considering minimizing language load; the format is not a barrier to student performance.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • Addressed issues of equity of utility for all populations. • Results of bias reviews and plans that have addressed any concerns. • At least two to three types of classification, reliability, and validity study data have been disaggregated by subgroups and meet the criteria. • Culturally diverse students were included throughout the entire process of test development. For example in the samples of pilot students, in cognitive interviews, etc. • The content of the reading materials does not favor mainstream culture. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>No bias reviews.</p> <p>Some culturally diversity mentioned, no further information on cognitive pieces of student abilities.</p> <p>Data is missing key components to make it a complete picture.</p> <p>No mention of a bias review Minimal subgroups represented</p>	<p>Does not meet: II</p> <p>Partially meets:</p> <p>Meets or exceeds:</p>

Third party evaluation conducted	Evidence reported to demonstrate that an independent, qualified third party has provided a thorough and unbiased evaluation of the quality of the assessment.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>Some evidence of teachers administering STEP in day to day classroom settings. Not scientifically as a study should be. See p. 60.</p> <p>Rasch review was noted</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>
Standardization of materials and procedures for administration	Administration protocol is scripted and provides precise guidelines; administration windows are clearly identified; materials are provided or clear guidelines are provided if materials are to be created; includes both electronic and hard copy administration manual that is clear and concise.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p>	Structure suggested and training information available.	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

		<p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		
Efficiency of administration	The amount of time needed to administer the assessment is reasonable and balanced to the information provided.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>
Efficiency of scoring	The amount of time needed to score the assessment is reasonable and balanced to	<p>DOES NOT MEET-evidence was not provided for</p>	Not clear	Does not meet:

	<p>the information provided; computer-assisted scoring is available; procedures for calculating scores are clear; scores can be stored and reported electronically.</p>	<p>this criteria or information does not demonstrate evidence.(0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Partially meets: II</p> <p>Meets or exceeds:</p>
<p>Accommodations clearly stated and described for students with disabilities and students with special needs (504, etc.)</p>	<p>The differing needs of students with disabilities are specifically addressed.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • Any accommodations do not compromise the interpretation or purpose of the test. • Specific administration guidelines are provided for implementing any accommodations. • How to address accommodations is specifically addressed in the training materials or program. • Suggested accommodations are research or evidence- 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided.</p>	<p>Available but not evidenced based.</p>	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

	based.	Information and data provided suggests acceptable or strong evidence. (2)		
Accommodations clearly stated and described for Second Language Learners	<p>The accommodations directly address the linguistic needs of the student.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> Any accommodation does not compromise the interpretation or purpose of the test. Specific administration guidelines are provided for implementing any accommodations. How to address accommodations is specifically addressed in the training. Suggested accommodations are research or evidence-based. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	A whole section provided.	<p>Does not meet:</p> <p>Partially meets: I</p> <p>Meets or exceeds: I</p>
Utility				
Scores are easily interpreted to determine a “significant reading deficiency”	<p>Scores clearly specify whether a student is categorized as having a “significant reading deficiency”.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> Score ranges or a scale is provided. Guides for 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p>PARTIALLY</p>	Unclear	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

	<p>interpretation of scores are provided.</p>	<p>MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		
<p>Cost effective: Materials, administration costs including personnel, scoring, and training</p>	<p>Materials are provided or easily accessible; time away from instruction is minimal; no additional personnel required; all costs inclusive including any additional data platform or storage costs; minimal data entry is required.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence.(0) PARTIALLY MEETS -partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or</p>	<p>No information provided No clear information on these indicators</p>	<p>Does not meet: II Partially meets: Meets or exceeds:</p>

		strong evidence. (2)		
Reports provide guidance for interpretation useful to educators, administrators, and parents	<p>Information is displayed in a format and language that is understandable to educators, administrators and parents;</p> <ul style="list-style-type: none"> • Data reports are easily read and interpreted. • Clear description of how to interpret results. • Reports provide trajectory for student progress. • District, school, classroom, and student reports provided. • Reports available in real-time. • Reports can be exported to data-base formats. • Reports available in languages other than English. • Customer service is available provided for users. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	Information is limited and useful	<p>Does not meet:</p> <p>Partially meets: II</p> <p>Meets or exceeds:</p>

Strengths:

- (1) Addresses the big picture of reading through the 5 components
- (2) Developmental
- (3) Five components addressed

Weaknesses:

- (1) Field testing
- (2) Validity of the test
- (3) Organization and access
- (4) Untimed

Recommendations:

Recommended _____

Not Recommended xx