

STAR Reading Spanish

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><b>Evidence includes:</b> 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"> <li>• Split-half reliability</li> <li>• Coefficient alpha</li> <li>• Test-retest reliability</li> <li>• Classification consistency</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –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>They provide evidence for split-half reliability with values of 0.85 – 0.90. They also provide evidence of generic reliability. They do not provide traditional internal consistency reliability coefficients because they are not calculated for adaptive tests</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>
	<p>Standard error of measurement or standard estimate of error is reported</p> <p><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• SEM estimates are reported for score</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate</p>	<p>They provide the SEM for score ranges, but they do not provide it for cut scores.</p>	<p>Does Not Meet –</p> <p>Partially Meets – 1</p> <p>Meets or</p>

	<p>ranges and cut-scores.</p> <ul style="list-style-type: none"> <li>SEM estimates are reported for score ranges and cut-scores for each assessment (grade-level, form, subtest).</li> </ul>	<p>evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS --</b> Information and data provided suggests acceptable or strong evidence. (2)</p>		Exceeds - I
	<p>Inter-rater reliability studies have been conducted. Study sample used to establish inter-rater reliability represents test administrators.</p> <p><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>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.</li> <li>Inter-rater reliability coefficients exceed .7.</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS --</b>most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence.</p>	<p>No information is provided regarding this, but it's computer adaptive so it's not expected.</p>	<p>Does Not Meet – I</p> <p>Partially Meets –</p> <p>Meets or Exceeds - I</p>

		(2)		
	<p>Studies have been conducted to establish reliability with all subcategories of students who will take the assessment.</p> <p><b>Evidence Includes:</b> 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><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>The pilot test does discuss the reliability results, but it does not disaggregate different subcategories of students who took the assessment.</p> <p>Lacking specific evidence</p>	<p>Does Not Meet –</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - 1</p>
<p>Alternative forms available for multiple assessments with demonstrated equivalence or comparability</p>	<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"> <li>• Technical reviews indicate all forms for each grade level have demonstrated evidence of comparability and</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data</p>	<p>Computer adaptive tests are their own alternate forms. Prior responses changes your future response.</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>

	<p>content specifications.</p> <p><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• Sufficient forms are provided to allow for progress monitoring between interim assessments.</li> <li>• Split-half reliability.</li> <li>• Coefficient alpha reliability.</li> </ul>	<p>provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –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><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• A clear description is provided that demonstrates the purpose of the assessment is to screen students for reading concerns.</li> <li>• Content specifications for each grade-level,</li> </ul>	<p><b>Rating</b></p> <p><b>Does Not Meet</b> – evidence was not provided for this criteria or information does not demonstrate evidence (0)</p> <p><b>Partially Meets</b> – partial evidence was provided related to the criterion and/or data provided</p>	<p>Page 26-34 of the technical manual, specifically Table 17 in the technical manual, provides a lot of information showing how the measures of a proficient reader in Spanish were considered specifically based on other Spanish reading assessment measures. Table 17 provides the evidence of concurrent validity. The correlations indicate that the tests arguably measure the same underlying reading ability between these assessments.</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>

	<p>including a complete description of the test content, purpose(s), and intended use(s), and assessment blueprint as appropriate, is provided.</p>	<p>demonstrates weak evidence. (1)</p> <p><b>Meets or Exceeds</b> – most information for the criterion is provided . Information and data provided suggests acceptable or strong evidence.</p>		
	<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><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• Field testing populations should be clear and should mirror the school/district demographics.</li> <li>• Statistics used to establish the reading levels are reported with both ELL and Non-ELL populations.</li> <li>• Findings from a content review by field experts, including teachers in tested</li> </ul>	<p><b>Meet</b> – evidence was not provided for this criteria or information does not demonstrate evidence (0)</p> <p><b>Partially Meets</b> – partial evidence was provided related to the criterion and/or data provided demonstrates</p>	<p>Reading level is reported as the level of achievement. The reading level was determined by many factors: the grade level placement of words on the Spanish vocabulary list and grade-leveled verb form scope and sequence, sentence difficulty progression, sentence length progression, use of graded target words, and the editorial judgment of U.S. and Mexico based editorial teams.</p> <p>Lacks study with evidence for multiple grades</p>	<p>Does Not Meet –</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - I</p>

	grade levels.	<p>weak evidence. (1)</p> <p><b>Meets or Exceeds</b> – most information for the criterion is provided . Information and data provided suggests acceptable or strong evidence.</p>		
	If appropriate, findings from alignment studies to demonstrate alignment with Colorado Academic Standards for Language Arts and resolution for any resulting concerns.	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	They state: “The assessment supports the Common Core State Standards and Colorado’s implementation of the READ Act.” While this is stated, there is no evidence provided to support this claim.	<p>Does Not Meet –</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - I</p>

	<p>There are studies of construct validity, such as convergent and discriminant analysis, demonstrating correlations of .7 or above.</p>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>They do not provide evidence of convergent or discriminant analysis, but they have a correlation between the skills rating and the scaled scores.</p> <p>Lacking in data for kindergarteners</p>	<p>Does Not Meet –</p> <p>Partially Meets – II</p> <p>Meets or Exceeds -</p>
<p>Evidence of criterion/predictive validity accurately identifying students with <i>“significant reading deficiency”</i></p>	<p>Evidence reported to 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><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• A clear definition of the criterion or measure that were used to establish concurrent validity.</li> <li>• Studies with similar assessments that demonstrate the</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR</b></p>	<p>Yes, they provide within-grade concurrent validity statistics. The statistics range from 0.5 – 0.89. They also state that this is an ongoing process.</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>

	assessment measures reading ability, not other irrelevant criteria. Predictive validity correlations above .7.	<b>EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)		
Determination of cut-scores based upon well-designed pilot study	<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><b>Evidence indicates:</b></p> <ul style="list-style-type: none"> <li>• Includes a description of the process used to establish the cut points.</li> <li>• A full description of the norming sample.</li> <li>• 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.</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and 2data provided suggests acceptable or strong evidence. (2)</p>	The cut scores are not provided in the technical manual	<p>Does Not Meet – I</p> <p>Partially Meets – I</p> <p>Meets or Exceeds -</p>
	Studies of classification accuracy analysis provide evidence that the measure appropriately identifies	<b>DOES NOT MEET</b> -evidence was not provided for this criteria or	They state: “the overall correlation was 0.37, indicating a weak relationship between	<p>Does Not Meet – I</p> <p>Partially</p>



	<p>students as indicated in the description of purpose of the assessment, demonstrating values that exceed .8 or higher.</p>	<p>information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)   <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>the computer adaptive STAR reading test scores and teachers’ ratings of their students’ reading skills”.</p>	<p>Meets – Meets or Exceeds - I</p>
	<p>Acceptable, recognized procedures are followed for setting cut-scores.</p>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided.</p>	<p>There is no mention of cut-scores, but they have scaled scores, the Spanish Instructional Reading level score, and growth scores to help educators to decide if students have an SRD.</p>	<p>Does Not Meet – Partially Meets – I Meets or Exceeds - I</p>

		Information and data provided suggests acceptable or strong evidence. (2)		
	SEM estimates are reported for cut-scores with guidance for score interpretation.	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence.(1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	There is no mention of cut scores, but they do give grade-level equivalencies for score interpretation.	<p>Does Not Meet –</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - I</p>
Universal Design	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	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was</p>	They state that “STAR is accessible to a wide diversity of learners”, because a computer adaptive test “adapts to the ability of the particular student.” However, they do not provide any evidence to support these claims.	<p>Does Not Meet – I</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - I</p>

	<p>performance.</p> <p><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• Addressed issues of equity of utility for all populations.</li> <li>• Results of bias reviews and plans that have addressed any concerns.</li> <li>• At least two to three types of classification, reliability, and validity study data have been disaggregated by subgroups and meet the criteria.</li> <li>• Culturally diverse students were included throughout the entire process of test development. For example in the samples of pilot students, in cognitive interviews, etc.</li> <li>• The content of the reading materials does not favor mainstream culture.</li> </ul>	<p>provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>No evidence provided for subgroups</p>	
<p>Third party evaluation conducted</p>	<p>Evidence reported to demonstrate that an independent, qualified third party has provided a thorough and unbiased evaluation of the quality of the assessment.</p>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates</p>	<p>They worked with Center for Applied Linguistics (CAL) and with Dr. Osuna specifically from the International Reading Association (IRA) to grade the Spanish vocabulary list.</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>

		<p>weak evidence. (1) <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		
<b>Administration &amp; Scoring</b>				
Standardization of materials and procedures for administration	<p>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>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>This test is similar to any other adaptive test and the instructions are standardized. Starting on p. 67 they show the standardized instructions and review the procedures for administration.</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>
Efficiency of	The amount of time	<b>DOES NOT</b>	The test only takes	Does Not

administration	needed to administer the assessment is reasonable and balanced to the information provided.	<p><b>MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	about 10 minutes and results are available immediately.	<p>Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>
Efficiency of scoring	The amount of time needed to score the assessment is reasonable and balanced to the information provided; computer-assisted scoring is available; procedures for calculating scores are clear; scores can be stored and reported electronically.		The scores are available immediately online. The reports provide data on students' performance in reading Spanish and teachers can drill down to view subgroups, classes, or individual students' level.	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>
Accommodations clearly stated and described for students with disabilities and	The differing needs of students with disabilities are specifically addressed.  <b>Evidence includes:</b>	<b>DOES NOT MEET</b> -evidence was not provided for this criteria or information	They state the allowable accommodations on p. 79 and 80 in detail.	<p>Does Not Meet –</p> <p>Partially Meets –</p>

<p>students with special needs (504, etc.)</p>	<ul style="list-style-type: none"> <li>Any accommodations do not compromise the interpretation or purpose of the test.</li> <li>Specific administration guidelines are provided for implementing any accommodations.</li> <li>How to address accommodations is specifically addressed in the training materials or program.</li> <li>Suggested accommodations are research or evidence-based.</li> </ul>	<p>does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>		<p>Meets or Exceeds - II</p>
<p>Accommodations clearly stated and described for Second Language Learners</p>	<p>The accommodations directly address the linguistic needs of the student.</p> <p><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>Any accommodation does not compromise the interpretation or purpose of the test.</li> <li>Specific administration guidelines are provided for implementing any accommodations.</li> <li>How to address accommodations is specifically addressed in the training.</li> <li>Suggested accommodations are research or evidence-</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and</p>	<p>The test is in students' native language</p>	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>

	based.	data provided suggests acceptable or strong evidence. (2)		
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><b>Evidence includes:</b></p> <ul style="list-style-type: none"> <li>• Score ranges or a scale is provided.</li> <li>• Guides for interpretation of scores are provided.</li> </ul>	<p><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p><b>PARTIALLY MEETS</b>-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	Score report is easy to read with three statistics explained in the report	<p>Does Not Meet –</p> <p>Partially Meets –</p> <p>Meets or Exceeds - II</p>
Cost effective: Materials, administration costs including personnel, scoring, and training	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><b>DOES NOT MEET</b>-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p><b>PARTIALLY MEETS</b> -partial evidence was</p>	On this website, <a href="http://www.renlearn.com/se/greatestvalue.aspx">http://www.renlearn.com/se/greatestvalue.aspx</a> they state that they test costs approximately \$10 per student.	<p>Does Not Meet –</p> <p>Partially Meets – I</p> <p>Meets or Exceeds - I</p>

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



	users.			
--	--------	--	--	--

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
Translation and adaptation procedure				
1. Translation has been provided by highly qualified personnel.	<p>Provide documentation on the translation team used to translate and adapt the test. Include the qualifications of the individuals who translated the test. The translation team should preferably include:</p> <ul style="list-style-type: none"> <li>translators who are native speakers in the target language</li> <li>specialists in reading in the target language</li> <li>bilingual educators (not to be confused with English as a Second Language (ESL) teachers or English as a Foreign Language (EFL) teachers or teachers of Spanish as a foreign language) in the target language.</li> </ul>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>It was not translated, it was written in Spanish. All writers are professional writers and editors who are native Spanish speakers and who have education backgrounds. They do not describe the kind of background in education they have, but they did address this section well.</p> <p>lacking in specific details about the creation of the test in Spanish</p>	<p>Does not meet –</p> <p>Partially Meets - II</p> <p>Meets or Exceeds-</p>
2. Pilot test sampling appropriately considers language diversity	<p>The translated test was piloted with a representative sample of speakers of the target language in the United States.</p>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests</p>	<p>There were three studies: a pilot study, a calibration study, and a research study. The pilot study included Spanish speaking students in grades 1-5 from South and West regions of the United States. There were 1,300 students in the pilot study. A Total of 12, 839 students participated in</p>	<p>Does not meet –</p> <p>Partially Meets - I</p> <p>Meets or Exceeds- I</p>

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
		acceptable or strong evidence. (2)	the calibration study.  Texas and California were the only 2 states cited, and the time frame is short	
<b>3. Consistency of appearance between the English language and the target language version of the test</b>	Formatting should remain consistent with the English language test version. Specifically, the font size of a translated test version should not be smaller than the English version. General ideas should be consistent with the English language test version.	<b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence. (0) <b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) <b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)	The two versions of the test are almost identical in appearance and are consistent in format between English and Spanish	Does not meet –  Partially Meets -  Meets or Exceeds- II
Criterion	Specific Indicators	Ratings		Notes
<b>Psychometric and measurement considerations:</b>				
<b>1. Construct validity for translated test versions</b>	Provide documentation to demonstrate that the test specifically identifies students with a “ <i>significant reading deficiency</i> ” in their native language. (i.e., test developers consider what constitutes a proficient reader in the target language rather than	<b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence.(0)  <b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)	Page 26-34 of the technical manual, provides a lot of information showing how the measures of a proficient reader in Spanish were considered specifically based on other Spanish reading assessment measures. Table	Does not meet –  Partially Meets - I  Meets or Exceeds- I

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
	<p>directly translating the measures of a proficient reader in English into the target language).</p> <p>Evidence is provided that the reading constructs measured by the test are relevant to the target language. As appropriate, information is reported on the procedures used to screen, select, and adapt the items of the test so that they are relevant and applicable to the target language.</p>	<p><b>MEETS OR EXCEEDS</b> – most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>17 provides the evidence of concurrent validity. The correlations indicate that the tests arguably measure the same underlying reading ability between these assessments.</p> <p>Evidence is insufficient. Evaluator is forced to make an assumption.</p>	
<p><b>3. Demonstrated comparability</b></p>	<p>Evidence is provided on the psychometric comparability of measures in English and measures in the target language.</p>	<p><b>DOES NOT MEET</b>- evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>- partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS</b> – most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>The sample score report defines the Independent Reading Level (IRL) for Spanish and for English in the same way. They were developed in different ways and do not explicitly compare the two tests. However, they provide this IRL measure, which shows the student’s grade level in reading for each language.</p> <p>Evidence is insufficient</p>	<p>Does not meet –</p> <p>Partially Meets - I</p> <p>Meets or Exceeds- I</p>
<p><b>4. Documentation on the interpretation of scores and the scaling of scores</b></p>	<p>Scaling information is provided to ensure appropriate interpretability of scores across language versions of the test so that educators and administrative officials</p>	<p><b>DOES NOT MEET</b>- evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS</b>- partial evidence was</p>	<p>Appropriate information is provided to the teacher and to the parents to interpret the Spanish language and English language</p>	<p>Does not meet –</p> <p>Partially Meets -</p> <p>Meets or</p>

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
	<p>know how to correctly interpret the scores obtained by the students in the translated version of the test.</p> <p>For example, do teachers need to scale the score of the translated test version in order to compare it with the English language version? If so, what kind of documentation is provided to assist teachers in this scaling process?</p>	<p>provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>versions of the test. The teachers do not need to scale the scores – the scores are provided in comparable forms already in the score report.</p>	Exceeds- II
<p><b>5. Evidence provided regarding investigation into potential item bias</b></p>	<p>Appropriate differential functioning items analyses across equivalent items have been conducted to examine bias for the same items across the two language versions. For example, for each item, is there a bias against students tested in the target language? Item bias reviews have been conducted and subsequent changes have been made based on recommendations.</p>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>The technical manual states that the test is free of bias and that it is fair; however, they do not provide any evidence to support this claim.</p> <p>Evidence is insufficient.</p> <p>Evaluator is forced to make assumptions.</p>	<p>Does not meet –</p> <p>Partially Meets - II</p> <p>Meets or Exceeds-</p>
Criterion	Specific Indicators	Ratings		Notes
<p><b>Equity and fairness considerations on the translated test version</b></p>				
<p><b>1. Consideration of appropriate dialect</b></p>	<p>The translation provides documentation to show</p>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria</p>	<p>No evidence provided regarding the</p>	<p>Does not meet – I</p>

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
	<p>that the translated test version does not privilege any dialect of the target language over others (e.g. Iberic Spanish - Spanish from Spain - is not privileged over Mexican or Puerto Rican dialects). Specifically, the translation procedures took into account the wide variety of dialects of the language speakers in the United States.</p>	<p>or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>dialect of the translated test version</p> <p>Evidence is insufficient.</p> <p>Evaluator is forced to make assumptions</p>	<p>Partially Meets - I</p> <p>Meets or Exceeds-</p>
<p><b>2. Appropriate cultural adaptation</b></p>	<p>Documentation is provided to show that items have been adapted to address cultural differences inherent to language. Cultural adaptations go beyond the superficial features of the contextual information provided by the items. For example, the items do not simply mention “Juan,” instead of “John,” as characters. Instead, consider how students’ experience may influence their interpretation of the items. Provide appropriate context for items to increase students’ access to the intended interpretation of the items.</p>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria or information does not demonstrate evidence. (0)  <b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)  <b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>The technical manual states that the test is free of bias and that it is fair; however, they do not provide any evidence to support this claim.</p> <p>Evidence is insufficient.</p> <p>Evaluator is forced to make assumptions.</p>	<p>Does not meet –</p> <p>Partially Meets - II</p> <p>Meets or Exceeds-</p>
<p><b>3. Address stereotypes</b></p>	<p>The cultural adaptation of the test is not based on stereotypes about</p>	<p><b>DOES NOT MEET-</b> evidence was not provided for this criteria</p>	<p>The Spanish technical manual states that no</p>	<p>Does not meet –</p>

Criterion	Specific Indicators	Ratings	Feedback from Reviewers	Tally of Rating
	cultures.	<p>or information does not demonstrate evidence. (0)</p> <p><b>PARTIALLY MEETS-</b> partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p><b>MEETS OR EXCEEDS –</b> most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>	<p>stereotypes are present, but they do not provide any evidence as to how they were able to ensure that there were no stereotypes made about different cultures.</p> <p>Evidence is insufficient.</p> <p>Evaluator is forced to make assumptions.</p>	<p>Partially Meets - II</p> <p>Meets or Exceeds-</p>

Strengths:

- 1) The test was not translated; it was developed in Spanish.
- 2) The test valued input from native Spanish speakers.
- 3) Reports, administration and scoring
- 4) Computer based is very student responsive with multiple versions and no issue of inter-rater reliability

Weaknesses:

- 1) The test was not accurate according to teachers' assessment of their students' reading skills, with a correlation of 0.37
- 2) There was no explicit mention of cut scores.
- 3) Lack for evidence and support for Spanish creation of the test.
- 4) Evidence presented started with children age 6, rather than kindergarten

Recommended: X X

Not Recommended: