Istation ISIP ER

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criterion | Specific Indicators | Rating | Feedback from Reviewers | Tally of rating |
| Validity, Reliability and Consistency in Scoring |  |  |  |  |
| Evidence of test reliability and consistency in scoring   | Results of reliability studies are reported for each grade assessment**Evidence includes:** The studies are appropriate given the purpose of the measure.For each grade-level, studies provide evidence of:* Split-half reliability
* Coefficient alpha
* Test-retest reliability
* Classification consistency
 | **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. Correlations demonstrate ranges of .7 or higher. (2) | No evidence of split-half reliability or coefficient alpha present.Page 41 Test is adaptive – students do not respond to a common set of items(excluding Coefficient alpha as a measure) Test-retest reliability – 0.927 to 0.970 (n=416) See table 7, p. 42 Test-retest completed across seven sections from Oct. to Feb. Internal Consistency high and stable over time(up to 5 months apart)  | Does not meet-Partially meets-1Meets or exceeds- 3 |
|  | Standard error of measurement or standard estimate of error is reported**Evidence includes:** * SEM estimates are reported for score ranges and cut-scores.
* SEM estimates are reported for score ranges and cut-scores for each assessment (grade-level, form, subtest).
 | **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 --**Information and data provided suggests acceptable or strong evidence. (2) |  Reviewers are not able to locate this information and the technical manual was downloaded from the internetNot reportedNo standard of error of measure included due to adaptability of assessment | Does not meet-4Partially meets-Meets or exceeds- |
|  | Inter-rater reliability studies have been conducted. Study sample used to establish inter-rater reliability represents test administrators. **Evidence includes:*** 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.
 | **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) | Computer administered assessmentComputer administered assessmentPage42States there was inter-rater reliability but does not provide data to correlate at grade level or specific coefficient. States IRR is within two points of all administrators. Due to adaptive, non-identical assessments, inter-rater reliability is difficult to determine.  | Does not meet-1Partially meets-Meets or exceeds-3 |
|  | Studies have been conducted to establish reliability with all subcategories of students who will take the assessment.**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. | **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) | Reliability was reported but not broken up by demographics groups.Page 38All subcategories of students are represented in the reliability and validity report Noted that studies were based off of one school district in North Texas.  | Does not meet-1Partially meets-Meets or exceeds-3 |
| Alternative forms available for multiple assessments with demonstrated equivalence or comparability | If alternative forms are provided, all forms have demonstrated evidence of equivalence or comparability such as test-retest, parallel form and internal consistency.* Technical reviews indicate all forms for each grade level have demonstrated evidence of comparability and content specifications.

**Evidence includes:*** Sufficient forms are provided to allow for progress monitoring between interim assessments.
* Split-half reliability.
* Coefficient alpha reliability.
 | **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 correlations demonstrate ranges of .7 or higher. (2) | Evidence of parallel forms, but not given about how often the test can be given. Computer based has bank of questions instead of alternate forms.This assessment is based on IRT not classical test theory, therefore it does not need alternative formsSince no student receives the same assessment, multiple forms are inherent in the program.Page 2 –states ISIP ER assess students automatically each month or on demands, as desired to measure progress throughout the school yearCoefficient alpha is noted as difficult to measure due to adaptability of test from student to student.Data shows comparability at each grade level to assessments | Does not meet-Partially meets-2Meets or exceeds-2 |
| Evidence of content and construct validity  | Evidence reported to demonstrate the assessment helps correctly identify students with *“significant reading deficiencies”* 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.**Evidence includes:*** 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 content, purpose(s), and intended use(s), and assessment blueprint as appropriate, is provided.
 | **Rating****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) | Statistics given about content validity of test.Page 1-web-based assessment of critical early reading skills that are predictive of later reading success Page 13 -Shows subtest administration for teach grade levelPage 14- Description of each subtest contains purpose and intended use of each. | Does not meet-Partially meets-Meets or exceeds-4 |
|  | Reading levels are reported for passages and how levels were established. Reading levels of assessment passages have been field-tested or have other evidence.**Evidence includes**:* 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.
 | **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 correlations demonstrate ranges of .7 or higher. (2) | Partial evidence showing they are reliable, does not show how they were made.Not evident in the technical manual or technical reportThe RFI and technical manual report that readability has been established using a “commonly accepted readability formula” but it does not say what this formula is. It also does not address whether these passages were field tested and which populations they were tested with. There is also no mention of a content review by a field of experts related specifically to the passages.Page18-To publisher, each of these passages was carefully written to conform to specific word level features follow linear story grammar structure and have readability according to a commonly accepted readability according to a commonly accepted readability formula for the end grade level in each grade.Demographics mirror school/districtLevels of assessment passages are not reportedPage 35- Field expert and  | Does not meet-2Partially meets-2Does not meet- |
|  | If appropriate, findings from alignment studies to demonstrate alignment with Colorado Academic Standards for Language Arts and resolution for any resulting concerns. | **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) | Alignment to both CAS are provided on the Istation web pageNo evidence of alignment to Colorado Academic Standards. | Does not meet-2Partially meets-Meets or exceeds-2 |
|  | There are studies of construct validity, such as convergent and discriminant analysis, demonstrating correlations of .7 or above. | **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) | Items previously tested for constructed validity were used to create. Author failed to demonstrate that this new test maintains construct validity.Not all correlations are at or about .70Page 42 Studies listed under validity section. However under multiple external measures, a correlation of .7 or higher is not always evident. | Does not meet-Partially meets- 3Meets or exceeds-1 |
| Evidence of criterion/predictive validity accurately identifying students with *“significant reading deficiency”*  | Evidence reported to demonstrate that the assessment has established criterion and/or predictive validity to correctly identify students with and without a *“significant reading deficiency.”****Evidence includes:**** 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.
 | **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) | Not all correlations are at or about .70Clear definition of criterion used to establish concurrent validity and the ranges of percentiles for the SRD are provided. However predictive validity is only included as a range of 0.59-0.82 on “a variety of reading test” with no explicit information provided. Criterion definitions use to establish validity for each measure compared page 40-41.Validity correlations of .7 or above are shown for significant number of external measures | Does not meet-Partially meets-2Meets or exceed-2 |
| Determination of cut-scores based upon well-designed pilot study  | The assessment has established cut-scores for decision making about students’ “*significant reading deficiency”* using adequate demographics representing (i.e., 10%ELL and 25% F/R lunch), appropriate criterion assessment, adequate sample size, and appropriate statistics.**Evidence indicates**: * Includes a description of the process used to establish the cut points.
* A full description of the norming sample.
* 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.
 | **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 2data provided suggests acceptable or strong evidence. (2) | Very clear percentiles.Three tiers of normative grouping based on percentile rank.Full description of norming sample which includes students from 24 states.Know demographics information is detailed.Large representative sample. | Does note meet-Partially meets-Meets or exceeds-4 |
|  | 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.  | **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) | Correlations, .7 .9Not clearly evident.No studies of classification accuracy are provided.Contains information on measures including criterion-referenced points, norm-referenced points and lexile scores to create an accurate analysis.Values are not shared to demonstrate values that exceed .8 | Does not meet-2Partially meets-1Meets or exceeds-1 |
|  | Acceptable, recognized procedures are followed for setting cut-scores. | **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) | Uses percentiles.Acceptable procedures for setting tiered cut scores based on percentile rank and use of standards deviations | Does not meet-Partially meets-Meets or exceed-4 |
|  | SEM estimates are reported for cut-scores with guidance for score interpretation. | **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) | Could not find.Standard error or measurement estimates are not evident to reviewers.No evidence of SEM estimates. | Does not meet-4Partially meets-Meets or exceeds- |
| 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 performance.**Evidence includes:** * 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.
 | **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) | Child frequently in game format, no evidence of bias reviews.Norming samples included a large percentage of culturally diverse students. However, statistics were not disaggregated by group to indicate that it is equally reliable and valid for all subgroups.Does not identify normed score based on disaggregated groups. Proof of bias control is not clear.Disaggregated subgroups include race, gender, and placement.Sample includes students from 24 states with appropriate representation; this qualifies as an adequately diverse population. | Does not meet-Partially meets-3Meets or exceeds-1 |
| 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. | **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) | Reviewers did not find evidence of this proposal.A panel of experts was used to review items for bias, however the overall assessment and study was not reviewed by an outside party.ISIP has documented independent reviews and studies from school districts comparing ISIP to state assessment and other widely accepted assessments.  | Does not meet-2Partially meets-Meets or exceeds-2 |
| 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. | **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) | Computer administered.Script provided in online teacher manual. Guidelines are provided for materials to be created. | Does not meet-Partially meets-Meets or exceeds-4 |
| Efficiency of administration  | The amount of time needed to administer the assessment is reasonable and balanced to the information provided. | **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) | Time to administer is around 30 minutes.Notice time limit for answering questions.Progress monitoring time shortened if only monitoring specific subtest... | Does not meet- Partially meets-Meets or exceeds-4 |
| 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. | **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) |  Immediate scoring as student takes assessment on computer.Efficient, storable computer-assisted. | Does not meet-Partially meets-Meets or exceed-4 |
| Accommodations clearly stated and described for students with disabilities and students with special needs (504, etc.) | The differing needs of students with disabilities are specifically addressed.**Evidence includes:*** 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-based.
 | **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) | Accommodation are addressed, however they are minimal at best and only provided for two populations of students.Supplemental materials contains adequate accommodations and modification for students with disabilities. | Does not meet-Partially meets-Meets or exceed-4 |
| Accommodations clearly stated and described for Second Language Learners  | The accommodations directly address the linguistic needs of the student.**Evidence includes**: * 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.
 | **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) | No accommodations listed for students other than a Spanish version of the test. Students speaking languages other than Spanish are not addressed.Specific guidelines for administering assessment with accommodations not located in technical manual or teacher guide.ELL accommodations given for lesson but not assessments. | Does not meet-1Partially Meets-1Meets or exceeds-2 |
| Scores are easily interpreted to determine a *“significant reading deficiency”*  | Scores clearly specify whether a student is categorized as having a *“significant reading deficiency”.* **Evidence includes:*** Score ranges or a scale is provided.
* Guides for interpretation of scores are provided.
 | **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) |  | Does not meet-Partially Meets-Meets or exceeds-2 |
| 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. | **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) | $5.50 per student | Does not meet-Partially meets-Meets or exceeds-3 |
| Reports provide guidance for interpretation useful to educators, administrators, and parents  | Information is displayed in a format and language that is understandable to educators, administrators and parents;* 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.
 | **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) |  Real time report, identify level of risk, shows growth, group in tiers are useful, results for classroom and by student.No districts reports provided | Does not meet-Partially meets-Meets or exceeds-3 |

**Strengths:**

1. Cost effective, can test whole class at once, screen for levels of risk. English version SRD is clear, reports are clear.
2. Computer based assessment and scoring streamlines administration and scoring procedures
3. An alignment with CAS and CCSS is provided on Istation web page.
4. Computer administration allows for standard administration and easy scoring.
5. Provides instructional materials to assist teachers in meeting student needs.

**Weaknesses:**

1. Not effective for monitoring.
2. Could not find SEM.
3. Statistical measurements such as SEM and construct validity are not readily evident.
4. Third party evaluation is not evident.
5. Not a stand-alone interim assessment. It is built to be used as an intervention curriculum with an assessment built in.
6. Requires school to have computers available for all students for testing purposes.

**Recommend: XXX Not recommend: XX**