AMIRA
LEARNING

Concurrent and Predictive Validity of ISIP Math with the Oklahoma State Testing Program

Mathematics, Grades 3 and 4 — Spring 2026

Prepared by: Amira Learning Research; Independent review was performed by ACS Ventures, LLC

Executive Summary

Amira Learning conducted a linking study of ISIP Math against the Oklahoma State Testing Program (OSTP) Mathematics assessment, using matched student records from the spring 2026 administration. The study covers 821 Grade 3 students and 830 Grade 4 students across four Oklahoma districts.

Concurrent validity. At Grade 3, ISIP Math administered in the spring benchmark window correlated .86 with the OSTP Mathematics Oklahoma Performance Index (95% CI .84–.88, n = 821). At Grade 4 the corresponding coefficient was .86 (95% CI .84–.88, n = 830).

Predictive validity. ISIP Math administered at the beginning of the year correlated .78 with Grade 3 OSTP Mathematics (95% CI .75–.81) and .76 at Grade 4. Mid-year administrations correlated .83 and .84 respectively.

Classification accuracy. Evaluated against the OSTP Basic-or-above standard, area under the curve ranged from .88 to .95 across grades and administration windows. The lower bound of the confidence interval exceeded .85 throughout. Five-fold cross-validated estimates matched in-sample estimates to within .002.

Score linkage. Equipercentile linking identifies the ISIP Math score corresponding to each Oklahoma performance standard. For Grade 3 in the spring window, the Basic standard fell

at an ISIP Math score of 492.0 — the 42nd percentile of the analysis sample — and the Proficient standard at 524.7, the 74th percentile. The proportion of students identified by each linked score fell within 1.1 percentage points of the proportion actually reaching the corresponding standard in every grade and window examined.

Summary of principal findings

Grade	Window	n	r with OSTP Math	95% CI	AUC
3	Beginning of year	788	.783	.754–.808	.883
3	Middle of year	800	.825	.801–.846	.921
3	End of year	821	.860	.841–.876	.951
4	Beginning of year	836	.761	.731–.788	.880
4	Middle of year	868	.838	.817–.857	.933
4	End of year	830	.860	.841–.877	.936

End-of-year administrations fall in the same testing window as OSTP and provide concurrent validity evidence; beginning- and middle-of-year administrations provide predictive validity evidence. AUC is evaluated against the OSTP Basic-or-above standard.

AMIRA
LEARNING

1. Purpose and Scope

1.1 Background

This study examines the relationship between ISIP Math and the Oklahoma State Testing Program (OSTP) Mathematics assessment, in order to establish how well a universal math screener administered during the school year corresponds to, and anticipates, student performance on Oklahoma's statewide summative assessment.

The question matters for practice. Oklahoma districts administer screeners three times a year to identify students who need additional support, and they are accountable for performance on a state assessment administered once each spring. If screener results correspond closely to state test performance, a district can act on screening information in the fall with reasonable confidence about where students stand relative to the standard they will ultimately be measured against.

1.2 Research questions

- How strongly does ISIP Math relate to OSTP Mathematics performance when the two are administered in the same testing window?
- How strongly does an ISIP Math score obtained earlier in the school year predict spring OSTP Mathematics performance?
- How accurately does ISIP Math classify students with respect to the OSTP Basic performance standard, and what cut scores support that classification?
- What ISIP Math score corresponds to the OSTP Proficient standard?

1.3 Grade scope

This study covers Grades 3 and 4. Grade 3 is the earliest grade at which OSTP is administered, making it the first point at which screener and state test results can be compared. Grade 4 is examined in the same districts and school year as a replication, providing evidence that any observed relationship is a property of the measures rather than of one grade's assessment or cohort.

AMIRA
LEARNING

2. Measures

2.1 ISIP Math

ISIP Math is a computer-adaptive mathematics assessment for students in prekindergarten through Grade 8, used for universal screening, benchmark assessment and continuous progress monitoring. Items are calibrated under a two-parameter logistic item response theory model, and the assessment adapts to each student's demonstrated ability during administration.

Districts administer ISIP Math benchmark assessments in three windows each school year: beginning of year (BOY), middle of year (MOY), and end of year (EOY). All three are analysed in this report.

The overall scale score

The measure analysed throughout this report is the ISIP Math overall scale score. Scores are reported on a single vertical scale spanning 100 to 900 across all grades, constructed by linking grade-specific item calibrations to a common metric. Because the scale is vertical, scores are directly comparable across grades and across administrations within a school year, which supports the measurement of growth as well as status.

ISIP Math is norm-referenced: a student's scale score is interpreted by comparison with a national norm sample of students at the same grade and time of year. Percentile ranks, five performance levels defined at the 20th, 40th, 60th and 80th percentiles, and three-level instructional tier groupings are reported alongside the scale score. Domain-level subscores are also produced; these are not analysed here.

Reliability

Reliability evidence for ISIP Math is reported in the ISIP Math Technical Report. Internal consistency reliabilities range from .90 to .95 across grades, and test–retest coefficients range from .76 to .84.

As a check specific to the present sample, marginal reliability of the overall score was also estimated from the conditional standard errors of measurement recorded for each assessment in the analysis sample, using the relationship between observed score variance and average conditional error variance. That estimate is .958, consistent with the published range.

AMIRA
LEARNING

2.2 OSTP Mathematics

The OSTP Mathematics assessment is Oklahoma's statewide summative assessment, administered once each spring in Grades 3 through 8. Results are reported on the Oklahoma Performance Index (OPI), together with one of four performance levels.

The OPI is a comparable scale spanning 200 to 399 on which 300 denotes Proficient in every grade. Performance-level boundaries below and above that point differ by grade. The spring 2026 online testing window for Grades 3–8 ran from 13 April to 13 May 2026.

Table 2.1 OSTP Mathematics performance levels and observed OPI ranges

Performance level	Grade 3 OPI	Grade 4 OPI	Interpretation
Below Basic	200–273	200–272	Has not performed at least at the Basic level
Basic	274–299	273–299	Partial mastery of essential grade-level knowledge and skills
Proficient	300–320	300–321	Mastery of grade-level subject matter; ready for the next grade
Advanced	321-399	322-399	Superior performance on challenging subject matter

Source: Oklahoma State Department of Education, OSTP and CCRA Performance Level Cut Points (2023). The Proficient threshold of 300 is fixed across grades and subjects by construction of the Oklahoma Performance Index.

The analysis uses the Basic threshold as the criterion for classification accuracy and cut-score derivation, and both the Basic and Proficient thresholds as targets for score linkage.

2.3 Reliability of the criterion measure

The precision with which any screener can predict OSTP performance is bounded by the reliability of OSTP itself. The Oklahoma State Department of Education reports total-test reliability and standard error of measurement in the annual OSTP technical report.

Table 2.2 Reliability of the OSTP Mathematics assessment

Grade	Students	Raw score max	Raw score mean	SD	Alpha	SEM
3	40,705	50	33.70	9.95	.92	2.78

Grade	Students	Raw score max	Raw score mean	SD	Alpha	SEM
4	39,524	50	31.34	10.27	.92	2.92

Source: Oklahoma State Department of Education, OSTP and CCRA Technical Report 2024–25, Table 8-1. Reliability is reported for the total test score, which is the criterion used throughout this study.

Internal consistency of .92 places OSTP Mathematics in the excellent range at both grades, comparable to the .90 to .95 range reported for ISIP Math (Section 2.1). Neither measure imposes a materially greater ceiling than the other on the observed relationship between them.

3. Sample and Data Sources

3.1 Participating districts

Four Oklahoma districts authorized matching of ISIP Math records to OSTP results for this study. Participation was voluntary. The sample is not a random or representative sample of Oklahoma districts, and this is addressed in Section 7.

3.2 Data sources

Amira benchmark records were supplied by each participating district as student-level exports. OSTP results were supplied by each district for the spring 2026 administration. Statewide and district-level OSTP performance figures used for comparison in Section 3.5 were obtained from published Oklahoma State Department of Education results files.

3.3 Analysis samples

Students with both a valid ISIP Math benchmark score and a valid OSTP Mathematics score in the same school year:

Table 3.1 Analysis sample sizes by grade and administration window

Grade	Beginning of year	Middle of year	End of year	Districts
3	788	800	821	4
4	836	868	830	4

3.4 Sample coverage and achievement relative to the state

Because no student demographic data was available, sample representativeness is characterised in two ways: the proportion of each district's tested population captured in the analysis, and the achievement profile of the participating districts relative to the state.

Coverage of the tested population

Table 3.2 Analysis sample as a proportion of each district's tested population

District	Grade 3 tested	Grade 3 analysed	%	Grade 4 tested	Grade 4 analysed	%
District A	28	28	100.0	36	35	97.2
District B	527	507	96.2	525	475	90.5
District C	179	138	77.1	207	156	75.4
District D	148	148	100.0	165	164	99.4
All districts	882	821	93.1	933	830	89.0

Tested counts are the valid OSTP Mathematics N published by OSDE for spring 2026.

The analysis therefore captures 93.1% of all Grade 3 students who took OSTP Mathematics in the participating districts, and 89.0% at Grade 4.

Achievement relative to the state

Table 3.3 OSTP Mathematics performance level distributions, spring 2026

Group	Gr.	Valid N	Mean OPI	Below Basic	Basic	Proficient or above
Oklahoma	3	47,970	286	32%	35%	32.6%
District A	3	28	286	—	54%	≥ 25.0%
District B	3	527	280	44%	30%	26.8%
District C	3	179	274	49%	30%	21.8%
District D	3	148	284	30%	43%	27.7%
Oklahoma	4	49,545	286	32%	35%	33.1%
District A	4	36	276	36%	53%	11.1%
District B	4	525	283	40%	32%	27.6%
District C	4	207	270	56%	32%	12.1%
District D	4	165	284	33%	36%	30.9%

Source: Oklahoma State Department of Education, Grades 3–8 OSTP District and School results, 2025–26. Proficient and Advanced are combined because OSDE suppresses one or both categories for several districts; where suppressed, the combined figure is obtained by subtracting the published Below Basic and Basic counts from the valid N. For District A at Grade 3, OSDE suppressed Below Basic as well, so the combined figure is a lower bound.

The participating districts span a range of achievement rather than clustering at one end. District D and B reach the Proficient standard at rates somewhat below the state figure at

both grades; District C is furthest below, with 49% of Grade 3 students scoring Below Basic against 32% statewide. In aggregate the sample is lower-achieving than the state: 26.3% of the Grade 3 analysis sample reached Proficient or above, against 33% statewide.

AMIRA
LEARNING

4. Methods

4.1 Record selection

A single completed benchmark result was retained per student per administration window. Progress-monitoring administrations were excluded. OSTP records coded as invalidated or as alternate assessments were excluded.

4.2 Validity coefficients

Pearson product-moment correlations were computed between the ISIP Math overall score and the OSTP OPI scale within each grade and administration window. Confidence intervals were derived by Fisher z transformation. Spearman rank-order correlations were computed alongside as a check on the influence of distributional shape.

Coefficients between the end-of-year administration and OSTP are reported as concurrent validity evidence, as the two fall in the same testing window. Coefficients between the beginning- and middle-of-year administrations and spring OSTP are reported as predictive validity evidence.

4.3 Classification accuracy

Classification accuracy was evaluated against the OSTP Basic-or-above standard. The event the screener is asked to detect is therefore a student scoring Below Basic; sensitivity is the proportion of such students correctly flagged, and specificity is the proportion of students reaching Basic or above who are correctly not flagged.

Area under the receiver operating characteristic curve was computed with DeLong standard errors, with confidence intervals constructed on the logit scale for improved small-sample coverage. Five-fold stratified cross-validated AUC was computed alongside each in-sample estimate as a guard against overfitting. Calibration was assessed by the Brier score.

4.4 Cut-score derivation

Candidate cut scores were derived from the ROC curve by three rules: the Youden J maximum; the most specific cut achieving sensitivity of at least .80; and the most specific cut achieving sensitivity of at least .90.

Sensitivity, specificity, positive and negative predictive value, overall and balanced accuracy, the Youden index, the flag rate and full confusion counts are reported at every candidate cut. Students scoring at or below the cut are flagged.

4.5 Score linkage

Single-group equipercentile linking was used to identify the ISIP Math score corresponding to a given OSTP performance standard, using mid-percentile ranks to handle ties, a three-point moving-average presmoother to reduce sampling irregularity, and enforced monotonicity.

Equipercentile results are a concordance function, not equating. The two instruments differ in construct coverage, format, blueprint and purpose. Linked scores support score interpretation and cut-score translation; they do not support score substitution, and a linked ISIP Math score should not be reported as though it were an OSTP score.

4.6 Software and reproducibility

Analyses were conducted in Python using pandas, NumPy, SciPy and scikit-learn. Cross-validation fold assignment uses a fixed random seed; no other stochastic component affects reported results. The analysis pipeline includes integrity assertions that halt execution if any district present in the input fails to appear in the analysis file, or if per-district counts fail to reconcile with pooled sample sizes.

AMIRA
LEARNING

5. Results

5.1 Descriptive statistics

Score distributions for the analysis samples. Statistics are computed on complete cases — students with both a valid ISIP Math score and a valid OSTP Mathematics score — so the sample sizes below are identical to those in the validity and classification tables that follow.

Table 5.1 ISIP Math overall score distributions by grade and window

Grade	Window	n	Mean	SD	Min	Median	Max
3	BOY	788	474.67	30.58	364.5	475.6	599.0
3	MOY	800	488.47	40.44	356.5	487.1	649.9
3	EOY	821	500.33	45.64	370.4	500.0	683.5
4	BOY	836	498.42	35.38	356.8	500.0	617.6
4	MOY	868	508.00	42.52	378.7	508.0	694.4
4	EOY	830	521.72	50.03	372.1	516.8	710.9

Scores rise across the three benchmark windows within each grade, consistent with growth over the school year. Score dispersion also widens across the year.

Table 5.2 OSTP Mathematics Oklahoma Performance Index distributions

Grade	Sample	n	Mean	SD	Min	Median	Max
3	BOY complete cases	788	280.67	27.27	218	280	399
3	MOY complete cases	800	280.26	27.54	218	280	399
3	EOY complete cases	821	280.29	27.40	218	280	399
4	BOY complete cases	836	280.99	27.33	221	280	399
4	MOY complete cases	868	280.78	27.35	221	280	399
4	EOY complete cases	830	280.58	27.36	221	279	399

OSTP is administered once each spring; each row shows the OPI distribution for the students analysed in that benchmark window. Proficient begins at OPI 300, so sample means fall approximately seven-tenths of a standard deviation below the Proficient standard.

5.2 Concurrent validity

ISIP Math administered in the spring benchmark window, in the same testing window as OSTP:

Table 5.3 Concurrent validity — end-of-year ISIP Math and spring OSTP Mathematics

Grade	n	r	95% CI	R^2	SEE (OPI)	p
3	821	.860	.841–.876	.739	14.01	< .001
4	830	.860	.841–.877	.740	13.97	< .001

SEE is the standard error of the estimate when OSTP OPI is predicted from the ISIP Math score.

5.3 Predictive validity

Table 5.4 Predictive validity — earlier ISIP Math administrations and spring OSTP Mathematics

Grade	Window	n	r	95% CI	R^2	SEE (OPI)
3	Beginning of year	788	.783	.754–.808	.612	16.99
3	Middle of year	800	.825	.801–.846	.680	15.58
4	Beginning of year	836	.761	.731–.788	.579	17.74
4	Middle of year	868	.838	.817–.857	.703	14.92

Coefficients strengthen as the interval to the state test shortens, from approximately .78 at the beginning of the year to .86 in the spring. This gradient partly reflects the widening score dispersion visible in Table 5.1: a narrower distribution mechanically attenuates a correlation. The comparatively small difference in classification accuracy across windows, reported in Section 5.4, indicates that the practical loss from screening early is smaller than the difference in correlations alone would suggest

5.4 Classification accuracy

Evaluated against the OSTP Basic-or-above standard. Base rate is the proportion of the sample scoring Below Basic — the group the screener is asked to identify.

Table 5.5 Classification accuracy against the OSTP Basic standard

Grade	Window	n	Base rate	AUC	95% CI	CV AUC
3	BOY	788	.407	.883	.857–.904	.883

Grade	Window	n	Base rate	AUC	95% CI	CV AUC
3	MOY	800	.412	.921	.901–.937	.922
3	EOY	821	.410	.951	.937–.963	.951
4	BOY	836	.404	.880	.855–.901	.881
4	MOY	868	.411	.933	.916–.947	.932
4	EOY	830	.417	.936	.919–.950	.936

CV AUC is the five-fold stratified cross-validated estimate. Agreement with the in-sample estimate to within .002 in every cell indicates the results are not products of overfitting.

5.5 Cut scores

Candidate ISIP Math cut scores for identifying students who will not reach the OSTP Basic standard. Students scoring at or below the cut are flagged.

Table 5.6 Cut scores, Grade 3

Window	Rule	Cut	Sens.	Spec.	PPV	NPV	Bal. acc.	Flag rate
BOY	Youden J maximum	469.8	.794	.844	.777	.857	.819	41.6%
BOY	Sensitivity ≥ .80	470.5	.801	.831	.765	.858	.816	42.6%
BOY	Sensitivity ≥ .90	480.2	.900	.675	.655	.908	.787	56.0%
MOY	Youden J maximum	483.3	.870	.819	.772	.900	.844	46.5%
MOY	Sensitivity ≥ .80	479.9	.809	.857	.799	.865	.833	41.7%
MOY	Sensitivity ≥ .90	488.5	.903	.755	.722	.917	.829	51.6%
EOY	Youden J maximum	487.2	.831	.934	.897	.888	.882	38.0%
EOY	Sensitivity ≥ .80	485.6	.810	.940	.904	.877	.875	36.8%
EOY	Sensitivity ≥ .90	498.4	.905	.804	.762	.924	.854	48.7%

Base rate .407 (BOY), .412 (MOY), .410 (EOY); n = 788 / 800 / 821.

Table 5.7 Cut scores, Grade 4

Window	Rule	Cut	Sens.	Spec.	PPV	NPV	Bal. acc.	Flag rate
BOY	Youden J maximum	499.1	.858	.767	.714	.888	.813	48.6%

Window	Rule	Cut	Sens.	Spec.	PPV	NPV	Bal. acc.	Flag rate
BOY	Sensitivity ≥ .80	496.0	.802	.797	.728	.856	.799	44.5%
BOY	Sensitivity ≥ .90	503.8	.902	.699	.670	.913	.801	54.4%
MOY	Youden J maximum	506.4	.899	.816	.773	.921	.858	47.8%
MOY	Sensitivity ≥ .80	496.9	.807	.898	.847	.869	.852	39.2%
MOY	Sensitivity ≥ .90	506.7	.902	.808	.767	.922	.855	48.4%
EOY	Youden J maximum	512.6	.879	.837	.794	.906	.858	46.1%
EOY	Sensitivity ≥ .80	507.5	.806	.882	.830	.864	.844	40.5%
EOY	Sensitivity ≥ .90	515.9	.902	.802	.765	.919	.852	49.2%

Base rate .404 (BOY), .411 (MOY), .417 (EOY); n = 836 / 868 / 830.

The sensitivity ≥ .80 rule reflects the NCII convention and is the recommended operational cut in each cell. At the Grade 3 end-of-year administration this cut achieves specificity of .940 and positive predictive value of .904 while flagging 36.8% of students — the most favourable operating point in the study.

Cut scores rise across the school year in both grades, tracking the growth visible in Table 5.1, so a single fixed cut is not appropriate across administration windows.

5.6 Score linkage to the OSTP performance standards

Equipercentile linking identifies the ISIP Math score that corresponds to a given point on the OSTP scale. Two standards are reported: the Basic standard, which is the criterion used for classification accuracy in Section 5.4 and therefore the standard relevant to screening decisions, and the Proficient standard at OPI 300, which is the standard used in state accountability reporting.

Table 5.8 ISIP Math scores linked to the OSTP Basic and Proficient standards

Gr.	Win.	n	Basic-linked score	Sample PR	Proficient-linked score	Sample PR
3	BOY	788	469.9	41.8	493.6	74.5
3	MOY	800	480.5	42.2	511.5	74.5
3	EOY	821	492.0	41.8	524.7	74.3
4	BOY	836	492.7	41.5	519.8	74.9
4	MOY	868	499.2	40.7	531.9	75.1
4	EOY	830	508.7	41.7	549.6	75.2

Sample PR is the percentile position of the linked score within the analysis sample; it is not a national percentile rank. Calibration: the Basic-linked score identified 57.8% to 59.3% of students as at or above the standard across the six cells, against 58.3% to 59.6% actually reaching Basic; the Proficient-linked score identified 24.8% to 25.7%, against 24.5% to 26.3% actually reaching Proficient. Agreement was within 1.1 percentage points in every cell and exact in one.

The percentile positions are notably stable: the Basic standard falls between the 41st and 42nd percentile, and the Proficient standard between the 74th and 75th, in every grade and administration window. That stability reflects the vertical construction of the ISIP Math scale, on which scores are comparable across grades and across administrations.

Convergence with the screening cut scores

The Basic-linked scores in Table 5.8 and the recommended screening cut scores in Section 5.5 are obtained by different procedures. The linked score comes from equipercentile concordance between two score distributions; the screening cut comes from optimising sensitivity and specificity on a receiver operating characteristic curve. Both are anchored to the same performance boundary in the same sample, so they are not statistically independent, but neither procedure constrains the other's result.

Table 5.9 Basic-linked score compared with the recommended screening cut score

Gr.	Win.	Basic-linked score	Screening cut (sens. ≥ .80)	Difference	Difference in SD
3	BOY	469.9	470.5	-0.6	0.02
3	MOY	480.5	479.9	+0.6	0.01
3	EOY	492.0	485.6	+6.4	0.14
4	BOY	492.7	496.0	-3.3	0.09

Gr.	Win.	Basic-linked score	Screening cut (sens. ≥ .80)	Difference	Difference in SD
4	MOY	499.2	496.9	+2.3	0.05
4	EOY	508.7	507.5	+1.2	0.02

Difference in SD expresses the gap relative to the standard deviation of the ISIP Math distribution in that grade and window.

The two procedures agree closely in all six cells. The largest difference is 6.4 scale points, or 0.14 standard deviations, and four of the six fall within 0.05 standard deviations. The sign of the difference varies across cells rather than running consistently in one direction, which is what would be expected if the remaining gap reflects sampling variation rather than a systematic difference between the methods.

The convergence supports the identified score region as a feature of the relationship between the two assessments rather than a product of a particular analytic choice, and it means that a district applying the recommended screening cut is applying a score that also corresponds closely to the OSTP Basic standard.

Interpreting the linked scores

The percentile columns make the two standards directly interpretable. The Basic standard falls at approximately the 41st percentile of the analysis sample — slightly more than half of students reached it. The Proficient standard falls near the 74th percentile. That difference in selectivity is a property of the Oklahoma standards rather than of the screener: statewide, 33% of Grade 3 and Grade 4 students reached Proficient in spring 2026, so Proficient is by construction a standard that a minority of students meet.

6. Interpretation

The measure is strongly associated with the state assessment

A concurrent coefficient of .86 at both grades indicates that ISIP Math and OSTP Mathematics measure substantially overlapping constructs, with approximately 74% of variance in state test performance accounted for.

Classification accuracy is high throughout

AUC between .88 and .95 across grades and windows, with confidence-interval lower bounds above .85 in every cell. At the Grade 3 spring administration, a cut meeting the .80 sensitivity standard also achieves specificity of .94, meaning the great majority of students who will reach the Basic standard are correctly not flagged.

Early screening remains informative

Beginning-of-year coefficients are lower than spring coefficients, but classification accuracy remains strong (AUC .88 at both grades), and the beginning-of-year administration leaves the full school year available for intervention.

Score linkage is well calibrated

Linked ISIP Math scores identify very nearly the same proportion of students as actually reach the corresponding standard — within 1.1 percentage points in every grade and window, for both the Basic and Proficient standards. The percentile position of each standard is stable across grades and administration windows, consistent with the vertical construction of the ISIP Math scale, and the score linked to the Basic standard agrees closely with the screening cut score derived from the ROC analysis. Together these support the use of ISIP Math benchmark scores to anticipate performance against Oklahoma's performance standards, within the interpretive limits set out in Section 4.5.

Precision of individual prediction

These results support screening and group-level inference. They are less precise for predicting an individual student's OPI score: the standard error of the estimate is approximately 14 OPI points at the spring administration, giving a 95% prediction interval of roughly ±27 points on a scale where Proficient begins at 300. ISIP Math should be used to identify students who require closer attention, not to forecast an individual score.

7. Limitations

Sample composition

The study draws on four districts that voluntarily authorised record matching. It is not a random or representative sample of Oklahoma districts. One district contributes approximately six in ten students in the analysis sample. In aggregate the sample is lower-achieving than the state, with 26.3% of Grade 3 students reaching Proficient against 33% statewide, although the participating districts span a range of more than 10 OPI points in mean performance (Section 3.5).

Scale properties

The OSTP OPI is a comparable rather than vertical scale. Within-grade coefficients are the primary results, and coefficients pooling grades are not reported as headline findings.

No subgroup analysis

Because no student demographic data was supplied, this study reports no evidence on whether classification accuracy differs across student groups. Statistical bias analysis is a separate standard under National Center on Intensive Intervention criteria and is not addressed here.

Study provenance

This analysis was conducted by Amira Learning. Independent review was performed by ACS Ventures, LLC. This analysis was conducted by Amira Learning. Independent review was performed by ACS Ventures, LLC. The review considered the linking study design and assumptions, sample sizes for the analyses, methods and processes, psychometric analysis and results, and technical documentation. In his opinion, dated 27 August 2026, the linking studies for both subject areas and grade levels were consistent with generally accepted testing industry practices with respect to evaluating concurrent and predictive evidence for uses of scores from interim or benchmark assessments. The review letter is available on request.

AMIRA
LEARNING

8. Conclusion

ISIP Math demonstrates strong concurrent validity with the Oklahoma State Testing Program Mathematics assessment at Grade 3 (r = .86) and Grade 4 (r = .86), across four Oklahoma districts and 1,651 matched student records from the spring 2026 administration.

Predictive validity from earlier administrations is also strong, at .78 to .84 depending on grade and window. Classification accuracy against the OSTP Basic standard exceeds National Center on Intensive Intervention criteria in every grade and administration window examined, reaching an area under the curve of .95 at Grade 3 in the spring window. Equipercentile linking places the ISIP Math score corresponding to the OSTP Proficient standard with close calibration to the observed proficiency rate.

9. Supporting Documentation

Supporting technical documentation is available on request, including the full analysis code, complete statistical output, and record-linkage and data-quality documentation. Amira Learning will provide these materials to the Oklahoma State Department of Education, the Office of Educational Quality and Accountability, or a designated reviewer.

10. References

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics, 44(3), 837–845.

Ketterlin-Geller, L. (2021). Istation's Indicators of Progress (ISIP) Math Technical Report. Dallas, TX: Istation.

Kolen, M. J., & Brennan, R. L. (2014). Test Equating, Scaling, and Linking: Methods and Practices (3rd ed.). New York: Springer.

National Center on Intensive Intervention. Academic Screening Tools Chart: Rating Rubric. Washington, DC: American Institutes for Research.

Oklahoma State Department of Education. *OSTP and CCRA Performance Level Cut Points (2023)*. Oklahoma City, OK.

Oklahoma Office of Educational Quality and Accountability. *Frequently Asked Questions*. Oklahoma City, OK.

Oklahoma State Department of Education. Grades 3–8 OSTP District and School Results (ELA/Math/Science), 2025–26. Oklahoma City, OK.

Oklahoma State Department of Education. OSTP and CCRA Performance Level Cut Points. Oklahoma City, OK.

Oklahoma State Department of Education. OSTP and CCRA Technical Report, 2024–25. Oklahoma City, OK.

Oklahoma State Department of Education. OSTP Grades 3–8 Key Dates, Spring 2026. Oklahoma City, OK.

Youden, W. J. (1950). Index for rating diagnostic tests. Cancer, 3(1), 32–35.