Concurrent Validity of Navvy Benchmark Scores: Correlations with AASA Summative Scores

Grades 4 and 7 Math, May 2025 Administration

Purpose

This brief report summarizes concurrent validity evidence for the Navvy Benchmark math assessments in grades 4 and 7. Navvy Benchmark percentage scores were correlated with scale scores from Arizona's Academic Standards Assessment (AASA) summative test, administered in May 2025, to evaluate the degree to which benchmark performance relates to end-of-year summative performance.

Method

Pearson product-moment correlations were computed separately for grade 4 (N = 429) and grade 7 (N = 470) between students' Navvy Benchmark scores (percentage correct) and their AASA math scale scores from the same academic year.

Results

Both grade levels showed strong, statistically significant positive correlations between Navvy Benchmark scores and AASA scale scores (see Table 1 and Figures 1–2).

The grade 4 correlation ($r = .866$, $p < .001$) was somewhat stronger than the grade 7 correlation ($r = .802$, $p < .001$), though both fall in the strong range.

Interpretation and Limitations

These correlations support the use of Navvy Benchmark scores as a general indicator of expected relative standing on the AASA summative assessment: students who score higher on the Navvy Benchmark tend to score higher on the AASA, and vice versa. However, several important limitations should be kept in mind when interpreting and communicating these results.

- **No score-level prediction.** A correlation coefficient describes the strength of a linear relationship across a group of students; it does not establish that a specific Navvy Benchmark score (e.g., 70%) predicts a specific AASA scale score or performance level for an individual student. Two students with the same benchmark score can land in different AASA performance categories.
- **Not a cut score or proficiency indicator.** These analyses do not establish, validate, or imply specific Navvy Benchmark thresholds that correspond to AASA proficiency levels (e.g., "Meets" or "Exceeds"). A separate standard-setting or linking study would be required to make claims of that kind.
- **Group-level, not individual-level, evidence.** The reported *r*, CI, and significance test describe the association across the full sample. They do not quantify the precision of any individual student's predicted AASA score and should not be used to generate individual score projections without additional analysis (e.g., regression with prediction intervals).
- **Explained variance is partial.** Even at $r = .866$, roughly 25% of the variance in AASA scores is unrelated to Navvy Benchmark performance (and roughly 36% at grade 7's $r = .802$). Other factors—test conditions, content differences, student growth between administrations, measurement error—account for the remainder.

- **Sample and timing specific.** Results reflect a single administration window (May 2025) and the students included in this sample. They should not be generalized to other grades, subjects, years, or populations without separate analysis.
- **Correlation, not causation.** The relationship is associational. It does not indicate that performance on one assessment causes performance on the other.

Summary

Navvy Benchmark scores show strong, statistically significant positive correlations with AASA summative scale scores at both grade 4 and grade 7 for the May 2025 administration. These findings support Navvy Benchmark performance as a general, group-level indicator of relative standing on the AASA. They do not support using a specific benchmark score to predict an individual student's expected AASA score or performance level; claims of that nature would require additional linking or standard-setting work.

Table 1

Grades 4 and 7 Correlational Study Results

Grade	N	*r*	95% CI	df	*t*	*p*
4	429	.866	[.840, .888]	427	35.74	< .001
7	470	.802	[.767, .832]	468	29.04	< .001

Figure 1

Grade 4: Navvy Benchmark Score (%) vs. AASA Scale Score (N = 429).

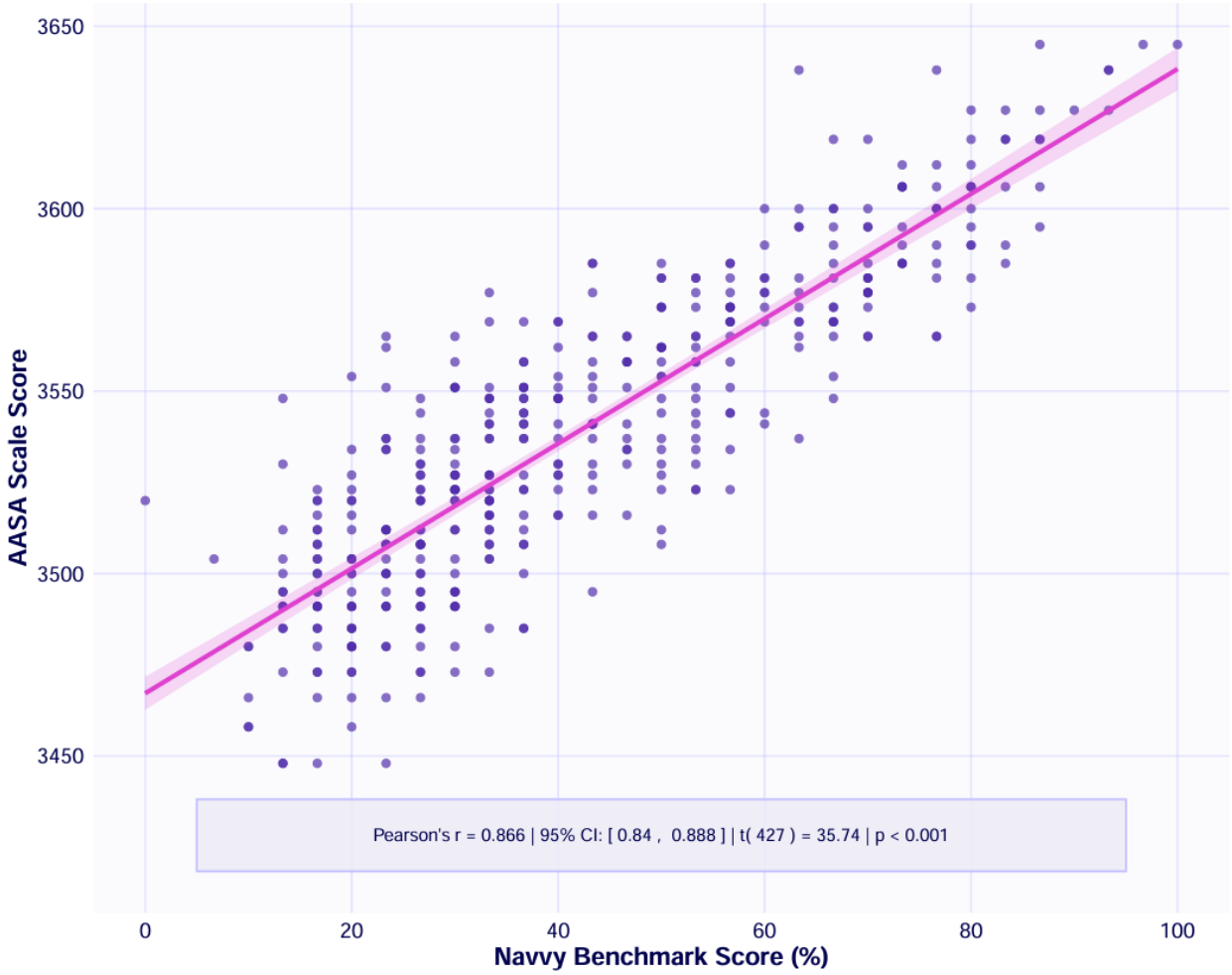


Figure 2

Grade 7: Navvy Benchmark Score (%) vs. AASA Scale Score (N = 470).

