



Assessment Manual

Melissa A. Gallagher

Published by Integrow™ Numeracy Solutions

<https://www.integrowmath.org/>

TABLE OF CONTENTS

1.0	Add+Vantage Math Screener Overview	4
1.1	Interpretation and Use Statements	4
1.2	Theoretical Framework	4
1.3	Standardized Administration Protocols.....	4
1.3.1	Format and Environment	4
1.3.2	Adaptive Administration and Probing	5
1.3.3	Documentation and Videotaping	5
1.4	Scoring and Coding.....	5
1.5	Assessment Schedule.....	5
1.6	Administrator Qualifications and Training Requirements.....	5
1.6.1	Required Professional Learning Tracks	6
1.6.2	Impact on Data Fidelity.....	6
2.0	Data Source and Sample Characteristics	7
2.1	Data Provenance.....	7
2.2	Analytic Sample Structure	7
2.3	Statistical Independence of Observations	7
2.4	Participant Demographics.....	9
3.0	Validity Evidence Related to Internal Structure	11
3.1	Model Specification and Linking	11
3.2	Sample Description	11
3.3	Model Fit and Dimensionality	11
3.4	Anchor Domain Discrimination.....	12
3.5	Differential Item Functioning (Grade-band non-invariance)	12
4.0	Scoring Methodology and Measurement Uncertainty.....	13
4.1	Latent Proficiency Estimation (θ)	13
4.2	Scoring Algorithm: Expected A Posteriori (EAP) Estimation.....	13
4.3	Score Precision and Conditional Standard Error.....	13
4.4	Scale-Level Precision and Information	13
5.0	Reliability and Internal Consistency	15
5.1	Standardized Training and Scoring Fidelity	15
5.2	Model-Based Marginal Reliability.....	15
5.2.1	Subgroup Reliability	15
5.3	Test Information and Measurement Precision.....	16

5.3.1	Grade-Band Targeting and Peak Information	17
5.3.2	Conditional Standard Error of Measurement (CSEM)	17
6.0	Reporting Scales and Normative Standing.....	19
6.1	Scale Score Transformation	19
6.2	Percentile Score Development and Reporting.....	19
6.2.1	Norm Development.....	19
6.2.2	Student-Level Percentile Reporting	20
6.2.3	Adaptive Banding in Grades 4 and 5	20
6.3	End-of-Year Domain Benchmarks	20
6.4	Rate of Improvement by Domain	21
6.5	Scale Score to Percentile Score Look Up Tables	24
7.0	Validity Evidence: Relation to External Variables.....	42
7.1	Data Linkage and Analytic Sample	42
7.2	Concurrent Validity Results.....	42
8.0	Classification Accuracy	43
8.1	Criterion and Risk Construct.....	43
8.2	Receiver Operating Characteristic (ROC) Analysis.....	43
8.2.1	Cut Score Decision Rule	43
8.3	Analytic Sample.....	43
8.4	Results: Spring Administration	43
8.5	Interpretation of Cut Scores.....	44
9.0	Evaluation of Bias: Concurrent Differential Prediction and Classification Fairness	46
9.1	Differential Prediction Framework (Logistic Interaction Models)	46
9.2	Results of the DTF.....	46
9.3	Interpretation	47
9.3.1	EL Status.....	47
9.3.2	Gender.....	48
9.3.3	Race.....	48
9.4	Overall.....	48
10.0	References.....	49

1.0 ADD+VANTAGE MATH SCREENER OVERVIEW

The Add+Vantage Math Screener is a diagnostic screening instrument designed to measure the numeracy development of students in Grades K–5. The Add+Vantage Math Screener evaluates student proficiency across seven developmental domains:

- Forward Number Word Sequence (FNWS)
- Backward Number Word Sequence (BNWS)
- Numeral Identification (NID)
- Structuring Numbers (SN)
- Addition and Subtraction (AS)
- Conceptual Place Value (CPV)
- Multiplication and Division (MD)

Each domain is evaluated on an ordinal scale (Levels 0–5) based on the Learning Framework in Number (LFIN). These levels are mathematically integrated to produce a composite scale score and percentile rank, facilitating both individual diagnostic profiling and aggregate screening.

1.1 Interpretation and Use Statements

Scores on the Add+Vantage Math Screener can be interpreted as the level of numeracy skill and understanding that a student has. Scores have two primary uses:

1. Screening: Identifying students performing significantly below grade-level expectations who may require intensive intervention (Tier 3).
2. Instructional Diagnostic: Providing teachers with a granular profile of a student's numerical strategies to inform differentiated Tier 1 and Tier 2 instruction.

The Add+Vantage Math Screener can be used as a universal screener in grades K-4, but is intended for supplemental use only in grade 5.

1.2 Theoretical Framework

The Add+Vantage Math Screener is grounded in the Learning Framework in Number (LFIN) (Wright & Ellemor-Collins, 2018). This framework conceptualizes mathematics acquisition not as the accumulation of isolated facts, but as a progression of increasingly sophisticated mental strategies. The screener identifies the "zone of proximal development" for each student within these developmental trajectories (Vygotsky, 1981).

1.3 Standardized Administration Protocols

To ensure the validity and reliability of the data presented in this manual, the Add+Vantage Math Screener must be administered according to the following standardized procedures.

1.3.1 Format and Environment

- Setting: Assessments are conducted in a one-on-one, clinical interview format in a quiet environment free from distractions.
- Positioning: The administrator should sit beside the student (rather than across) to maintain a clear line of sight to the student's non-verbal strategies (e.g., finger counting, subitizing movements).

- Materials: Standardized kits, including counters and screens, must be organized prior to administration to maintain the flow of the assessment.

1.3.2 Adaptive Administration and Probing

The Add+Vantage Math Screener utilizes an adaptive testing logic. While items are standardized, the entry and exit points within a domain are determined by the student's performance on previous tasks.

- Probing: Administrators must use standardized, non-leading probes (e.g., "How did you work that out?") to identify the student's underlying strategy.
- Neutrality: Administrators must refrain from providing verbal or non-verbal feedback regarding the correctness of a response to prevent the inflation of scores through "teaching-to-the-test" during the interview.

1.3.3 Documentation and Videotaping

While digital or paper-based coding is required, videotaping is recommended as a standard for high-stakes screening or diagnostic precision. Videotaping allows for a retrospective analysis of subtle student behaviors and non-verbal strategy use that may be missed during live administration.

1.4 Scoring and Coding

The coding process involves mapping observed student behaviors and strategies to the ordinal levels of the LFIN.

- Standardized Coding: Each task is coded for both accuracy and the strategy utilized.
- Level Determination: A student's level within a domain is determined by the highest strategy consistently demonstrated, rather than a simple sum of correct answers.
- Timing: To minimize recall bias, coding should be completed immediately following the interview or during a review of the video record.

1.5 Assessment Schedule

To support a Multi-Tiered System of Supports (MTSS), the following administration window is recommended:

- Universal Screening (Fall): Establishes baseline data and identifies students for intervention.
- Progress Updates (Winter): Minimally three updates throughout the year to monitor growth and adjust instructional groups.
- Summative Evaluation (Spring): Evaluates end-of-year proficiency and informs placement for the subsequent academic year.

1.6 Administrator Qualifications and Training Requirements

To maintain the psychometric integrity and standardized administration of the Add+Vantage Math Screener, all administrators must complete rigorous professional learning. This ensures that the qualitative coding of student strategies is consistent across diverse educational settings.

1.6.1 Required Professional Learning Tracks

The Add+Vantage Math Screener training is divided into two comprehensive courses, totaling approximately 60 hours of professional development. The requirements are stratified by grade level to reflect the complexity of the domains administered:

- Primary Grades (K–1): Administrators must complete Add+Vantage Course 1. This course focuses on the foundational domains of Forward and Backward Number Word Sequences, Numeral Identification, Structuring Numbers, and Addition and Subtraction.
- Intermediate Grades (2–5): Administrators must complete both Add+Vantage Course 1 and Course 2. Course 2 provides the specialized training required to assess and code higher-level multiplicative reasoning (Multiplication and Division) and Conceptual Place Value.

1.6.2 Impact on Data Fidelity

This extensive training requirement is a prerequisite for valid score interpretation. The training ensures that administrators can correctly identify the threshold behaviors that distinguish one LFIN Level from another, maintain affective neutrality during the interview process, and minimize inter-rater variance by adhering to the standardized coding rubrics established by the Learning Framework in Number.

2.0 DATA SOURCE AND SAMPLE CHARACTERISTICS

This section details the participant sample used to establish the normative benchmarks, reliability estimates, and validity evidence for the Add+Vantage Math Screener.

2.1 Data Provenance

The evidence presented in this manual is derived from a longitudinal dataset of a large Midwestern school district spanning six academic years: 2018–2019 through 2023–2024. The district serves a diverse student population, providing a robust localized sample for evaluating the technical adequacy of the Add+Vantage Math Screener in a real-world Multi-Tiered System of Supports context.

2.2 Analytic Sample Structure

The dataset consists of 42,377 assessment records corresponding to 8,628 unique students (see Table 1). Because the dataset is longitudinal, some students contributed records in more than one academic year as they progressed across grade levels. As a result, the number of assessment records exceeds the number of unique students. For descriptive summaries of unique students, each student was counted once based on the most recent year of data collection. This convention was used only for person-level tabulation and helps explain the comparatively large number of unique students shown in Grade 5. Unless otherwise noted, psychometric analyses were conducted using the full set of assessment records retained for calibration.

Table 1

Unique Student Counts and Assessment Records by Grade

Grade	Unique Students	Total Assessment Records
Kindergarten	923	6,911
1	960	7,220
2	956	7,341
3	941	7,256
4	994	7,007
5	3,854	6,642
Total	8,628	42,377

2.3 Statistical Independence of Observations

For the purposes of this manual, analyses were conducted using the following independence rules to ensure psychometric integrity:

- Classification Accuracy and Reliability: These analyses were stratified by grade. Because the Add+Vantage Math Screener is a grade-normed instrument, a student’s performance in Grade 1 (Year 1) is treated as a distinct measurement event from their performance in Grade 2 (Year 2). This approach ensures that the "Risk" definition remains calibrated to the specific developmental expectations of each grade level.
- Normative Percentiles: Percentile ranks were calculated based on the distribution within each grade-year cohort to account for any year-over-year shifts in district performance.

- Demographic Representation: To prevent the over-representation of long-term residents in demographic summaries, all percentage-based subgroup distributions (e.g., Race, Gender, ELL status) are calculated based on the unique student count (see Table 3
- Table 5).

2.4 Participant Demographics

Table 2

Demographic Composition of the Full Student Sample in the Classification and Concurrent Validity Analyses

Grade	Total	Sex		Race/Ethnicity							Language	
		Female	Male	American Indian or Alaska Native	Asian	Black or African American	Hispanic/Latino	Native Hawaiian or Other Pacific Islander	Two or more races	White	ELL	
1	653	310	343	2	51	64	58	0	66	412	121	
2	595	298	297	2	44	60	50	0	64	375	104	
3	1,189	582	607	1	95	97	98	0	124	774	223	
4	1,123	521	602	3	95	87	101	0	128	709	228	
5	1,115	517	598	3	93	93	87	1	122	716	202	
Total	4,675	2,228	2,447	11	378	401	394	1	504	2,986	878	

Table 3

Demographic Composition of the Unique Student Sample in the Classification and Concurrent Validity Analyses

Grade	Total	Sex		Race/Ethnicity							Language	
		Female	Male	American Indian or Alaska Native	Asian	Black or African American	Hispanic/Latino	Native Hawaiian or Other Pacific Islander	Two or more races	White	ELL	
1	651	310	341	2	51	64	58	0	66	410	120	
2	587	295	292	2	44	58	50	0	63	370	96	
3	686	344	342	1	57	69	53	0	64	442	110	
4	643	303	340	1	48	48	68	0	71	407	136	
5	1,115	517	598	3	93	93	87	1	122	716	202	
Total	3,682	1,769	1,913	9	293	332	316	1	386	2,345	664	

Table 4*Demographic Composition of the Full Student Sample in IRT Analysis*

Grade	Total	Sex			Race/Ethnicity							
		Female	Male	Transgender	American Indian or Alaska Native	Asian	Black or African American	Hispanic/Latino	Native Hawaiian or Other Pacific Islander	Two or more races	White	NA
KG	6,911	3,303	3,608	0	18	555	558	582	3	633	4,562	0
1	7,220	3,540	3,680	0	25	572	567	588	0	657	4,811	0
2	7,341	3,595	3,746	0	17	594	554	540	5	697	4,934	0
3	7,256	3,508	3,744	4	15	629	525	548	1	679	4,857	2
4	7,007	3,327	3,678	2	13	627	487	543	5	647	4,685	0
5	6,642	3,195	3,445	2	27	581	493	503	5	588	4,445	0
Total	42,377	20,468	21,901	8	115	3,558	3,184	3,304	19	3,901	28,294	2

Table 5*Demographic Composition of the Unique Student Sample in IRT Analysis*

Grade	Total	Sex			Race/Ethnicity							
		Female	Male	Transgender	American Indian or Alaska Native	Asian	Black or African American	Hispanic/Latino	Native Hawaiian or Other Pacific Islander	Two or more races	White	NA
KG	923	423	500	0	6	72	110	78	1	83	573	0
1	960	470	490	0	6	70	107	93	0	86	598	0
2	956	488	468	0	6	58	121	99	0	89	583	0
3	941	475	466	0	2	76	109	69	0	85	599	1
4	994	452	542	2	2	71	92	97	3	98	633	0
5	3,854	1,846	2,008	1	16	331	288	292	3	325	2,600	0
Total	8,628	4,154	4,474	3	38	678	827	728	7	766	5,586	1

3.0 VALIDITY EVIDENCE RELATED TO INTERNAL STRUCTURE

To establish a unified reporting scale across varying grade levels, the Add+Vantage Math Screener utilized a Multiple-Group Graded Response Model (MG-GRM). This approach accounts for planned missingness in the grade-adaptive administration design while ensuring score comparability.

3.1 Model Specification and Linking

The calibration was conducted across four distinct grade bands (kindergarten, grade 1, grades 2–3, and grades 4–5). To establish a common metric (vertical scaling), two anchor domains administered across all levels—Structuring Numbers and Addition and Subtraction—were constrained to be invariant across groups. Latent means and variances were freely estimated for each grade band to reflect developmental growth.

3.2 Sample Description

The sample is approximately balanced by gender, with 48% female and 52% male students (see Table 4). This distribution remains consistent across all grade levels, ensuring that the instrument's norms are not biased toward a specific gender group. The student population is predominantly White (67%), with significant representation from diverse racial and ethnic subgroups. Notably, students identifying as Black or African American comprise 8% of the sample, followed by Asian (8%), Hispanic/Latino (8%), and students of two or more races (9%).

3.3 Model Fit and Dimensionality

A restricted anchors-only multiple-group model using SN and AS was examined as a supplementary fit-evaluation model, because these domains were administered across grade bands and provided the most stable common anchor set. MD and PV were excluded from the fit-evaluation model because they were not administered in K–1 and exhibited category sparsity that complicated stable fit estimation, though they were retained in the operational scoring model. FNWS, BNWS, and NID were also excluded from the fit-evaluation model because cross-grade comparisons indicated substantial noninvariance and limited utility as stable anchors, particularly in grades 4 and 5. The final model was selected as the best-fitting unidimensional multiple-group graded response model based on likelihood-based fit comparisons, information criteria, nested multiple-group invariance comparisons, and residual diagnostics. Limited-information global fit indices such as RMSEA, CFI, TLI, and SRMSR were not available for the final model because the multiple-group graded-response structure, partial invariance specification, and sparse category usage within some grade-band subgroups prevented stable computation of M2/C2-based omnibus fit statistics. Taken together, the results support a practically unidimensional model for the core numeracy construct, with SN and AS serving as the primary cross-grade anchors and other domains incorporated with group-specific flexibility in the operational model.

Separate one-factor graded response models fit within each grade band showed uniformly high factor loadings (approximately 0.79 to 0.96) and high proportions of explained variance (> 0.60), indicating a strong dominant latent dimension within bands. In particular, SN and AS loaded strongly across all grade bands, supporting their role as the most stable cross-grade anchors in the final multiple-group model.

3.4 Anchor Domain Discrimination

To support the vertical scale, two anchor domains (SN and AS) were constrained to be invariant across grade bands (K–1, 2–3, and 4–5). The discrimination parameters (a) for these domains were high, indicating a strong relationship between the tasks and the latent proficiency construct (see Table 6), and showed appropriately ordered category thresholds across the latent continuum, supporting their use as stable cross-grade anchors in the final multiple-group graded response model.

Table 6

Anchor Domain Discrimination Parameters (a)

Domain	Discrimination (a)	Interpretation
Addition and Subtraction (AS)	3.32	Strong
Structuring Numbers (SN)	3.09	Strong

Note: Parameters constrained invariant across all grade bands.

The invariance of these anchor parameters helps ensure that cross-grade differences on the common scale are referenced to stable linking domains, while allowing other domains to vary across grade bands when warranted by the data.

In the final multiple-group graded response model, Grade 1 was used as the reference group, and estimated latent means increased monotonically across grade bands from KG to Grades 4–5. The anchor domains SN and AS were constrained to have identical discrimination and threshold parameters across groups, thereby defining the common vertical scale, while group means and variances were freely estimated.

3.5 Differential Item Functioning (Grade-band non-invariance)

To evaluate cross-grade comparability, a sequence of nested multiple-group graded response models was fit to test whether domain parameters could be held invariant across grade bands. The analysis began with a model constraining all domains to function equivalently across groups, followed by a series of partially invariant alternatives in which selected domains were freed across grade bands while invariant anchor domains maintained the common metric. Results showed that constraining MD and PV to invariance produced substantially worse fit than allowing them to vary across groups. Subsequent comparisons demonstrated that FNWS, BNWS, and NID also exhibited meaningful grade-band non-invariance. Across the full sequence of comparisons, both likelihood-ratio tests and information criteria favored increasingly flexible partially invariant models.

The final selected model was the best-fitting unidimensional multiple-group graded response model and retained only SN and AS as invariant anchors across grade bands. Parameters for BNWS, FNWS, NID, MD, and PV were freely estimated by group in the operational calibration. These findings indicate that a fully invariant unidimensional model was not tenable, but that a defensible partially invariant unidimensional vertical scale could be established using SN and AS as the common linking domains.

4.0 SCORING METHODOLOGY AND MEASUREMENT UNCERTAINTY

This section describes the procedures used to derive individual student scores and the associated estimates of measurement error.

4.1 Latent Proficiency Estimation (θ)

Student performance is reported as an estimate of the latent numeracy trait (θ) obtained from the final multiple-group graded response model (MG-GRM). This model places students from all grade bands on a common vertically linked scale while allowing selected non-anchor domains to vary across groups when warranted by the calibration results. As a result, reported scores are comparable across grade bands within the common vertical metric defined by the invariant anchor domains.

4.2 Scoring Algorithm: Expected A Posteriori (EAP) Estimation

Individual student proficiency estimates were obtained using expected a posteriori (EAP) scoring. Under EAP scoring, a student's estimate is the mean of the posterior distribution of proficiency conditional on the student's observed response pattern, the calibrated item parameters from the final operational model, and the group-specific distribution implied by the multiple-group calibration.

EAP estimation was used because it yields stable model-based proficiency estimates and reduces the influence of extreme response patterns, particularly at the tails of the score distribution. This property makes EAP scoring well suited for screening applications, where consistent and interpretable score reporting is important.

4.3 Score Precision and Conditional Standard Error

The Add+Vantage Math Screener reports a person-specific conditional standard error for each score. This value is obtained from the posterior standard error associated with the EAP estimate. Because the model is estimated within an IRT framework, measurement precision is conditional rather than constant across the score scale.

A smaller standard error indicates greater precision in the student's estimated proficiency, whereas a larger standard error indicates greater uncertainty. These standard errors depend on the amount of information provided by the student's response pattern and the degree to which that response pattern is targeted by the calibrated domains. Accordingly, scores located near the center of the scale are typically estimated more precisely than scores at the extreme low or high ends of the continuum.

In the transformed reporting metric, score precision is expressed on the same scale as the reported score by applying the linear score-transformation constant to the posterior standard error of θ . This ensures that reported uncertainty remains directly aligned with the model-based precision of each individual estimate.

4.4 Scale-Level Precision and Information

Beyond individual SEMs, the precision of the Add+Vantage Math Screener is evaluated at the aggregate scale level using Test Information Functions (TIF). Test information is calculated across a proficiency continuum ranging from -4.0 to +4.0.

The relationship between information (I) and the conditional standard error (SE) is defined by the following function:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Standard error curves are derived for each grade band to identify where the instrument provides the highest degree of reliability. As documented in Section 0, the Add+Vantage Math Screener is designed to minimize measurement error in the "at-risk" regions of the scale, ensuring that the classification of students for intensive intervention is based on the most precise data points available.

5.0 RELIABILITY AND INTERNAL CONSISTENCY

5.1 Standardized Training and Scoring Fidelity

To ensure high levels of consistency, all administrators must complete an intensive 60-hour standardized training protocol. This program is specifically designed to mitigate the risk that raters might differ in their consistency or their overall scoring level. The training includes:

- **Expert Modeling:** Teachers watch and analyze multiple videos of experts administering the tool to establish a clear "gold standard" for scoring.
- **Fidelity of Administration:** Teachers are required to video record their own administration of each domain. This is followed by a structured self-reflection on their recording to ensure they are adhering to the established interview protocols.
- **Practice Scoring:** Teachers engage in rigorous practice scoring using expert-validated videos. This process ensures that different educators, when presented with the same student performance, arrive at the same score, directly addressing the core requirements of inter-rater consistency.

The Add+Vantage system replaces "ad hoc" item-based measures with a comprehensive training and certification process that ensures the scores yielded across different educators and time periods remain consistent and reliable for making high-stakes instructional decisions.

5.2 Model-Based Marginal Reliability

Marginal reliability ($\bar{\rho}$) was used to summarize the consistency of model-based scores in the analytic sample. In the IRT framework, this statistic compares the average posterior error variance of the score estimates with the observed variance of the estimated latent trait in the sample. It was computed as:

$$\bar{\rho} \approx 1 - \frac{(\frac{1}{N}) \sum_{i=1}^N SE_{\hat{\theta}_i}^2}{var(\hat{\theta})}$$

where $SE_{\hat{\theta}_i}$ = the posterior standard error of examinee i 's EAP θ estimate.

For the full analytic sample, the marginal reliability of the Add+Vantage Math Screener was 0.916 (95% bootstrap CI [0.915, 0.917]), indicating strong overall score consistency for the vertically linked scale.

5.2.1 Subgroup Reliability

Reliability was evaluated at the grade level to ensure technical adequacy for grade-specific screening, with 95% bootstrap confidence intervals within each grade (see Table 7).

Table 7*Marginal Reliability of the Add+Vantage Math Screener by Grade*

Grade	N	[95% CI]
Kindergarten	6911	0.90 [0.90, 0.91]
1	7220	0.88 [0.87, 0.88]
2	7341	0.84 [0.83, 0.85]
3	7256	0.82 [0.81, 0.83]
4	7007	0.71 [0.70, 0.72]
5	6642	0.57 [0.55, 0.59]
Overall (K-5)	42,377	0.92 [0.91, 0.92]

Marginal reliability was highest in the earlier grades and declined across the upper grades, with the lowest value observed in Grade 5. This pattern is consistent with the conditional standard error and test information results, which indicated that the vertical scale is less well targeted to the upper end of the proficiency continuum. Because marginal reliability is sample dependent, the higher overall value for the pooled K–5 sample should not be interpreted as applying equally to each grade. The overall estimate reflects the larger combined variance across grades, whereas the grade-specific estimates provide the more relevant evidence for within-grade screening use.

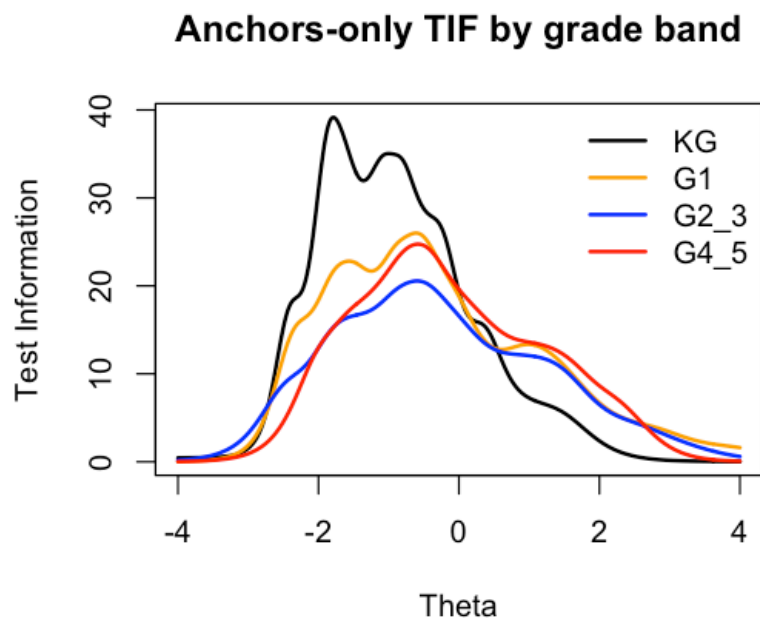
The comparatively lower reliability in Grade 5 suggests that the current item pool provides less precise measurement for students at the upper end of the scale. Accordingly, score interpretation appears strongest for Grades K–4, while Grade 5 scores should be interpreted with greater caution and may be most appropriate when used in conjunction with other sources of information.

5.3 Test Information and Measurement Precision

The reliability of the Add+Vantage Math Screener is further evaluated through the Test Information Function (TIF), which illustrates how measurement precision varies across the latent proficiency continuum (θ , see Figure 1). Higher values of test information indicate greater measurement precision and lower conditional standard error of measurement. Unlike a single classical reliability coefficient, the TIF identifies the portions of the proficiency scale at which the assessment is most and least informative.

Figure 1

Anchors-Only Test Information Function (TIF) by Grade Band



5.3.1 Grade-Band Targeting and Peak Information

The TIFs shown in Figure 1 reflect the information provided by the two invariant anchor domains, Structuring Numbers (SN) and Addition and Subtraction (AS), across grade bands. These anchors provided substantial measurement precision in all groups, with the highest information concentrated at lower theta levels for Kindergarten and Grade 1, and progressively broader, right-shifted information for Grades 2–3 and Grades 4–5. This pattern is consistent with the developmental ordering observed in the estimated latent means, with younger students located lower and older students higher on the common vertical scale.

Importantly, these curves represent the precision of the anchor set, not the full operational model. As such, they are most appropriately interpreted as evidence that the common linking domains function effectively across grade bands and provide a stable measurement backbone for the vertical scale. Taken together, the anchors-only TIF results support the use of SN and AS as strong and developmentally appropriate cross-grade linking domains.

5.3.2 Conditional Standard Error of Measurement (CSEM)

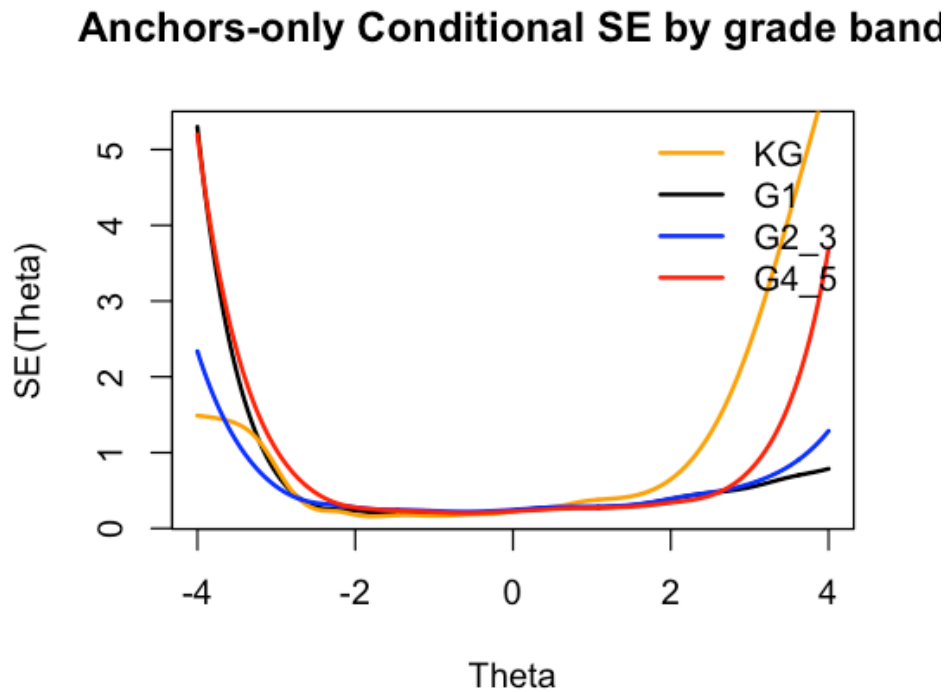
Measurement precision is inversely related to the Conditional Standard Error of Measurement (CSEM). Consequently, the regions of peak information correspond to the lowest levels of measurement error.

The anchors-only conditional standard error functions showed that measurement error was lowest across the central portion of the latent proficiency continuum and increased at the extreme low and high ends of the scale (see Figure 2). This pattern was consistent across grade bands and indicates that the anchor domains, SN and AS, provide the most precise cross-grade linking

information in the range where the groups overlap most strongly, while precision is reduced at the tails of the continuum.

Figure 2

Anchors-Only Conditional Standard Error of Measurement by Grade Band



6.0 REPORTING SCALES AND NORMATIVE STANDING

6.1 Scale Score Transformation

Individual latent proficiency estimates (θ) were obtained using expected a posteriori (EAP) scoring from the final multiple-group graded response model. These latent scores were then transformed to the operational reporting scale using a linear transformation:

$$\text{ScaleScore} = A \cdot \theta + B$$

This transformation preserved the interval-scale properties of the latent metric while expressing scores on a reporting scale designed for operational use. The constants A and B were derived from the distribution of theta estimates in the analytic sample to produce a scale-score mean of 200 and standard deviation of 40 (see Table 8).

Table 8

Transformation Constants and Latent Distribution

Parameter	Value	Description
A (slope)	23.351	The scaling factor (multiplier) for the θ estimate.
B (intercept)	173.399	The constant added to center the reporting scale.
Mean(θ)	1.139	The average latent proficiency of the calibration sample.
SD(θ)	1.713	The standard deviation of latent proficiency in the sample.

The transformed scores had a mean of 200 and a standard deviation of 40 in the scoring sample, confirming that the reporting metric was applied as intended. For reporting purposes, scale scores were rounded to the nearest integer.

To report score precision on the same metric, the reporting-scale conditional standard error was computed by applying the slope constant to the posterior standard error of the latent estimate:

$$\text{ScaleSE} = |A| \cdot \theta_{se}$$

Score precision varied across examinees, with a median reporting-scale standard error of 7.21 points and an interquartile range from 5.65 to 12.92 points, reflecting the conditional nature of measurement precision under the IRT model.

6.2 Percentile Score Development and Reporting

Add+Vantage Math Screener percentile scores were developed to support interpretation of student performance within a grade and administration window. The final methodology was designed to maximize interpretive usefulness while limiting false precision in score ranges affected by sparse samples or ceiling effects.

6.2.1 Norm Development

Percentile cut scores were defined separately by grade and administration window - fall, winter, and spring. Winter and spring cut scores were based directly on observed empirical score distributions.

Fall cut scores were developed differently. Because fall samples were comparatively sparse and potentially selective, fall norms were constructed using a developmental smoothing approach

rather than relying solely on the raw fall distributions. For kindergarten, fall cut scores were anchored to the observed kindergarten winter distribution. For Grades 1-5, each fall percentile cut score was positioned between the previous grade spring cut score and the same grade winter cut score. This approach produced fall norms that were more stable, developmentally ordered, and distinct from winter norms.

The final percentile cut-score table therefore combined constructed fall cut scores with observed winter and spring cut scores.

6.2.2 *Student-Level Percentile Reporting*

Student percentile results were assigned by matching each student's grade, window, and Add+Vantage Math Screener scale score to the final cut-score table and interpolating within the corresponding grade-window distribution (see Section 6.5 for the full scale score to percentiles tables).

Reporting rules varied by grade to reflect the precision supported by the data. Kindergarten through Grade 2 and Grade 3 fall used exact percentile reporting. Grade 3 winter and spring used exact percentiles through the 90th percentile, with higher scores reported as "90th percentile and above."

For Grades 4 and 5, exact percentile reporting was used only below the 40th percentile. Above that point, students were assigned to percentile bands rather than exact percentile ranks.

6.2.3 *Adaptive Banding in Grades 4 and 5*

In the upper grades, observed score distributions showed substantial ceiling effects, especially in winter and spring. In some grade-window combinations, the cut scores corresponding to the 60th and 80th percentiles were identical, and in the most compressed cases the 40th, 60th, and 80th percentile cut scores all collapsed to the same scale score.

To avoid reporting distinctions not supported by the data, upper-band reporting in Grades 4 and 5 was made adaptive to the observed cut-score structure. When the upper cut scores were distinct, three bands were used. When the 60th- and 80th-percentile cut scores were tied, the upper two bands were collapsed. When the 40th-, 60th-, and 80th-percentile cut scores were all tied, all upper distinctions were collapsed into a single category: 40th percentile and above.

6.3 End-of-Year Domain Benchmarks

Domain-level end-of-year benchmarks were established for each Add+Vantage Math Screener domain (FNWS, BNWS, NID, SN, AS, MD, and PV) to provide grade-specific expectations for student performance at the conclusion of the school year (see Table 9). Because the Add+Vantage Math Screener is administered adaptively, with certain domains introduced only when students demonstrate readiness for the underlying mathematical content, the benchmark methodology was designed to respect the assessment's adaptive structure while producing interpretable, norm-referenced expectations.

Benchmarks were derived from end-of-year (spring) administrations using the full set of available student records. For each grade and domain, the benchmark cut score was operationalized as the highest integer score that at least sixty percent of administered students scored at or above. This rule produces a benchmark that reflects the level of performance

attained by a clear majority of students who reached the domain by end of year and is therefore interpretable as a typical end-of-grade expectation for students for whom the domain is developmentally appropriate.

To account for the adaptive administration of the assessment, an explicit coverage rule was applied prior to reporting any benchmark. For each grade and domain, the proportion of students with an administered score on that domain at end of year was computed relative to the total number of students assessed at that grade. A benchmark was reported only when at least sixty percent of students at a given grade were administered the domain. Grade-by-domain combinations falling below this coverage threshold were not assigned a benchmark and are explicitly indicated as not reported, accompanied by the corresponding administration rate. This convention prevents the publication of benchmarks based on small or non-representative subsets of students who advanced to a domain before the majority of their peers, and clearly distinguishes domains that are developmentally appropriate at a given grade from those that are typically introduced in later grades.

Each benchmark therefore reflects two conditions: that the domain was administered to a substantial majority of students at the grade level, and that, within that administered sample, the cut score corresponds to a level reached by at least sixty percent of students. Benchmarks are reported as integer scores aligned to the domain's native scale, supporting direct interpretation by educators without requiring conversion through percentile or scaled-score systems.

The resulting benchmark table provides grade-specific expectations for each domain that is routinely administered at that grade. This offers educators a transparent and defensible reference for evaluating end-of-year student performance on the Add+Vantage Math Screener while preserving the integrity of the screener's adaptive design.

Table 9

End of Year Benchmarks by Grade Level and Domain

Grade	AS	BNWS	FNWS	MD	NID	PV	SN
KG	1	3	4	NR	2	NR	0
1	3	5	5	NR	4	NR	2
2	4	5	5	2	4	2	4
3	5	5	5	3	5	4	5
4	5	5	5	4	5	4	5
5	5	5	5	5	5	5	5

Note. NR = Not Reported. Less than 60% of students in these grade levels were assessed on these domains.

6.4 Rate of Improvement by Domain

To characterize short-term growth patterns on the Add+Vantage Math Screener domain scores, we estimated student-level Rate of Improvement (ROI) for each Add+Vantage domain and summarized ROI by grade level. Because several Add+Vantage domains have limited score ranges, we evaluated the extent to which ceiling effects influenced observable growth and computed ROI estimates restricted to students with room to grow on each domain (see Table 10).

The seven Add+Vantage domains analyzed were Forward Number Word Sequence (FNWS), Backward Number Word Sequence (BNWS), Numeral Identification (NID), Place Value (PV), Multiplication and Division (MD), Addition and Subtraction (AS), and Structuring Numbers (SN). Each domain score was paired with its domain-specific administration date. ROI was computed at the level of student $\times$ school year $\times$ grade $\times$ domain. For each unit of analysis, the first and last valid administration dates were identified, along with the corresponding first and last domain scores, and the elapsed time between the two administrations was computed in days. ROI was then calculated as the simple first-to-last slope—that is, the difference between the last and first scores divided by the number of elapsed days between them—and was rescaled to a weekly metric by multiplying the daily rate by seven.

To ensure that ROI estimates reflected meaningful change rather than short-interval retest variability, ROI was computed only for student-domain records that included at least two observations, that spanned a minimum of 14 days between the first and last administrations, and that had beginning of year (BOY) scores below the ceiling score of 5. Records that did not meet these criteria were excluded from grade-level ROI summaries. Several Add+Vantage domains exhibit substantial ceiling effects, particularly at the upper elementary grades, where many students begin the school year already at the domain maximum. In such cases, grade-level ROI estimates computed across the full sample can be attenuated, masking meaningful growth that occurs among students who began the year below the ceiling. To address this, we computed for each grade $\times$ domain combination the percentage of students who began the year at the domain ceiling and then estimated ROI restricted to students with room to grow on that domain (defined as a first score below the domain ceiling).

For each grade $\times$ domain combination, the primary reported growth statistic was the mean weekly ROI among students with room to grow, reported alongside the percentage of students starting at the ceiling. Reporting these two values together provides an interpretable estimate of typical weekly improvement for students who could realistically demonstrate growth on a given domain, while also documenting the extent to which ceiling effects constrain observable growth at each grade level.

Table 10

Rate of Improvement and Percent of Students Scoring at Ceiling at the Beginning of the School Year by Grade and Domain

Grade	Domain	N	ROI	Percent at Ceiling at BOY
KG	AS	2829	0.029	0%
KG	BNWS	2671	0.064	6%
KG	FNWS	2388	0.075	16%
KG	NID	2812	0.055	1%
KG	SN	2859	0.030	0%
1	AS	2812	0.046	2%
1	BNWS	1619	0.062	43%
1	FNWS	1060	0.063	63%
1	MD	217	0.026	0%
1	NID	2561	0.042	8%
1	PV	222	0.036	1%
1	SN	2770	0.058	3%
2	AS	2153	0.044	22%
2	BNWS	674	0.056	75%
2	FNWS	363	0.049	87%
2	MD	1953	0.034	1%
2	NID	1755	0.037	34%
2	PV	1995	0.052	3%
2	SN	2109	0.058	24%
3	AS	1265	0.035	53%
3	BNWS	273	0.047	90%
3	FNWS	159	0.047	94%
3	MD	2262	0.051	11%
3	NID	653	0.037	73%
3	PV	1833	0.042	26%
3	SN	1162	0.041	57%
4	AS	698	0.031	73%
4	BNWS	132	0.049	95%
4	FNWS	65	0.041	97%
4	MD	1356	0.034	47%
4	NID	254	0.036	90%
4	PV	1344	0.035	48%
4	SN	716	0.042	72%
5	AS	370	0.036	85%
5	BNWS	55	0.048	98%
5	FNWS	36	0.049	98%
5	MD	807	0.033	66%
5	NID	116	0.028	95%
5	PV	893	0.032	63%
5	SN	408	0.044	83%

6.5 Scale Score to Percentile Score Look Up Tables

Table 11

Kindergarten Fall

Scale Score	Percentile
<=116	1
117-118	2
119	3
120-121	4
122	5
123	6
124	8
125	10
126	11
127	12
128-129	13
130	14
131	15
132	16
133	17
134-135	18
136	19
137	21
138	25
139	30
140	40
141	48
142	60
143	63
144	67
145	70
146	74
147	76
148	78
149	80
150	82
151	84
152	85
153	87
154	88
155-156	90
157-158	91

159-160	92
161-162	93
163-165	94
166-167	95
168-169	96
170-171	97
172-173	98
>=174	99

Table 12*Kindergarten Winter*

Scale Score	Percentile
<=114	1
115-116	2
117-118	3
119-121	4
122	5
123	6
124	8
125	10
126	12
127	13
128	15
129	17
130	18
131	20
132	22
133	24
134	26
135	28
136	30
137	34
138	37
139	40
140	43
141	46
142	49
143	52
144	57
145	61
146	64
147	68
148	71
149	74
150	77
151	78
152	80
153	82
154	83
155	84
156	86

157	87
158	89
159	90
160	91
161	92
162	93
163	94
164-165	95
166-167	96
168-169	97
170-172	98
>=173	99

Table 13*Kindergarten Spring*

Scale Score	Percentile
<=114	1
115-118	2
119-122	3
123-126	4
127-129	5
130	6
131-132	7
133	8
134	9
135	10
136	11
137	12
138	14
139	15
140	16
141	17
142	18
143	20
144	22
145	24
146	26
147	28
148	30
149	33
150	35
151	37
152	40
153	43
154	45
155	48
156	51
157	54
158	57
159	60
160	63
161	66
162	69
163	72
164	75

165	77
166	79
167	81
168	83
169	85
170	86
171	88
172	90
173	91
174	93
175	94
176-177	95
178-179	96
180-181	97
182-183	98
>=184	99

Table 14*Grade 1 Fall*

Scale Score	Percentile
<=117	1
118-121	2
122-124	3
125-128	4
129-131	5
132-133	6
134	7
135	8
136-137	9
138	10
139	11
140	12
141	14
142	15
143	16
144	17
145	18
146	20
147	22
148	23
149	25
150	27
151	29
152	31
153	33
154	35
155	37
156	40
157	42
158	44
159	46
160	48
161	51
162	53
163	56
164	59
165	61
166	64
167	67

168	69
169	72
170	74
171	77
172	79
173	81
174	83
175	84
176	86
177	88
178	89
179	91
180	92
181	93
182	94
183-184	95
185-186	96
187-189	97
190-192	98
>=193	99

Table 15*Grade 1 Winter*

Scale Score	Percentile
<=120	1
121-124	2
125-128	3
129-131	4
132-134	5
135-136	6
137-138	7
139	8
140-141	9
142	10
143	11
144	12
145	13
146	14
147	15
148	17
149	18
150	19
151	21
152	22
153	24
154	25
155	27
156	28
157	29
158	31
159	33
160	35
161	37
162	39
163	41
164	43
165	44
166	46
167	48
168	50
169	52
170	54
171	56

172	59
173	61
174	63
175	65
176	68
177	70
178	72
179	74
180	77
181	79
182	81
183	82
184	84
185	86
186	87
187	89
188	90
189	91
190	92
191	93
192	94
193-195	95
196-198	96
199-202	97
203-205	98
>=206	99

Table 16*Grade 1 Spring*

Scale Score	Percentile
<=128	1
129-132	2
133-137	3
138-141	4
142-144	5
145	6
146-147	7
148-149	8
150-151	9
152	10
153-154	11
155	12
156	13
157	14
158	15
159	16
160-161	17
162	18
163	19
164	20
165	21
166	22
167	24
168	25
169	26
170	28
171	30
172	31
173	33
174	35
175	36
176	38
177	40
178	42
179	44
180	46
181	48
182	49
183	52

184	54
185	56
186	58
187	60
188	63
189	66
190	69
191	71
192	74
193	76
194	79
195	80
196	82
197	83
198	84
199	85
200	86
201	87
202	88
203	89
204	90
205	91
206-207	92
208	93
209	94
210-211	95
212-213	96
214-215	97
216-217	98
>=218	99

Table 17*Grade 2 Fall*

Scale Score	Percentile
<=130	1
131-134	2
135-139	3
140-143	4
144-146	5
147-148	6
149-150	7
151-152	8
153-154	9
155	10
156	11
157-158	12
159	13
160	14
161	15
162	16
163-164	17
165	18
166	19
167	20
168	21
169	22
170	24
171	25
172	26
173	28
174	30
175	31
176	33
177	35
178	36
179	38
180	40
181	42
182	44
183	46
184	48
185	50
186	52

187	54
188	56
189	58
190	60
191	63
192	65
193	67
194	69
195	71
196	73
197	75
198	77
199	79
200	80
201	82
202	83
203	84
204	85
205	86
206	88
207	89
208	90
209-210	91
211	92
212-213	93
214	94
215-216	95
217-218	96
219-220	97
221-223	98
>=224	99

Table 18*Grade 2 Winter*

Scale Score	Percentile
<=132	1
133-137	2
138-142	3
143-147	4
148-151	5
152-153	6
154	7
155-156	8
157-158	9
159	10
160-161	11
162	12
163	13
164	14
165-166	15
167	16
168	17
169	18
170-171	19
172	20
173	22
174	23
175	24
176	26
177	27
178	29
179	30
180	32
181	34
182	36
183	37
184	39
185	41
186	43
187	45
188	47
189	49
190	51
191	53

192	55
193	57
194	59
195	61
196	63
197	64
198	66
199	68
200	69
201	71
202	72
203	74
204	75
205	77
206	78
207	80
208	81
209	82
210	84
211	85
212	86
213	88
214	89
215	90
216	91
217-218	92
219-220	93
221	94
222-223	95
224-226	96
227-229	97
230-231	98
>=232	99

Table 19*Grade 2 Spring*

Scale Score	Percentile
<=138	1
139-143	2
144-148	3
149-154	4
155-157	5
158-160	6
161-162	7
163-165	8
166-167	9
168-169	10
170	11
171-172	12
173	13
174	14
175	15
176-177	16
178	17
179	18
180	19
181	20
182	21
183	22
184	23
185	25
186	26
187	27
188	28
189	29
190	31
191	32
192	34
193	35
194	37
195	38
196	39
197	41
198	43
199	44
200	46

201	48
202	49
203	51
204	53
205	54
206	56
207	58
208	60
209	61
210	63
211	65
212	66
213	68
214	70
215	71
216	73
217	74
218	76
219	77
220	78
221	79
222	80
223	82
224	83
225	85
226	87
227	89
228	90
229-231	91
232-233	92
234-235	93
236-237	94
238-240	95
241-244	96
245-248	97
249-251	98
>=252	99

Table 20*Grade 3 Fall*

Scale Score	Percentile
<=141	1
142-146	2
147-151	3
152-157	4
158-160	5
161-163	6
164-165	7
166-168	8
169-170	9
171-172	10
173	11
174-175	12
176	13
177	14
178	15
179-180	16
181	17
182	18
183	19
184	20
185	21
186	22
187	23
188	24
189	26
190	27
191	28
192	29
193	30
194	31
195	33
196	34
197	35
198	37
199	38
200	40
201	41
202	43
203	44

204	46
205	47
206	49
207	50
208	52
209	54
210	55
211	57
212	59
213	61
214	62
215	64
216	66
217	67
218	69
219	71
220	72
221	74
222	76
223	77
224	78
225	79
226	80
227	81
228	83
229	85
230	86
231	88
232	89
233	90
234-136	91
237-238	92
239-241	93
242	94
242-243	94
244-245	95
246-248	96
249-250	97
251-252	98
>=253	99

Table 21*Grade 3 Winter*

Scale Score	Percentile
<=146	1
147-151	2
152-156	3
157-161	4
162-165	5
166-167	6
168-170	7
171-172	8
173-174	9
175-176	10
177-178	11
179	12
180	13
181	14
182-183	15
184	16
185	17
186	18
187-188	19
189	20
190	21
191	23
192	24
193	25
194	26
195	27
196	28
197	29
198	30
199	31
200	32
201	34
202	35
203	36
204	37

205	38
206	40
207	41
208	42
209	44
210	45
211	47
212	48
213	50
214	51
215	53
216	55
217	56
218	58
219	60
220	61
221	63
222	65
223	67
224	68
225	70
226	72
227	74
228	75
229	76
230	77
231	79
232	80
233	81
234	82
235	84
236	85
237	87
238	88
239	90
>=240	90th percentile and above

Table 22*Grade 3 Spring*

Scale Score	Percentile
<=146	1
147-153	2
154-159	3
160-166	4
167-170	5
171-173	6
174-175	7
176-178	8
179-181	9
182-183	10
184	11
185	12
186-187	13
188	14
189-190	15
191	16
192-193	17
194	18
195	19
196-197	20
198	21
199	22
200	23
201	24
202	25
203	26
204	27
205	28
206	29
207	30
208	31
209	32
210	33
211	34
212	35
213	36
214	37
215	38
216	39

217	40
218	41
219	42
220	44
221	45
222	46
223	48
224	49
225	51
226	52
227	53
228	54
229	56
230	57
231	58
232	59
233	60
234	61
235	62
236	63
237	63
238	64
239-240	65
241	66
242-243	67
244	68
245	69
246-247	70
248	71
249-250	72
251	73
252-253	74
>=254	90th percentile and above

Table 23*Grade 4 Fall*

Scale Score	Percentile
<=150	1
151-156	2
157-162	3
163-168	4
169-172	5
173-175	6
176-177	7
178-180	8
181-182	9
183-184	10
185	11
186-187	12
188	13
189-190	14
191	15
192-193	16
194	17
195-196	18
197	19
198	20
199	21
200	22
201	23
202	24
203	25
204	26
205	27
206	28
207	29
208	30
209	31
210	32
211	33
212	34
213	35
214	36
215	37
216	38
217	39

218	40
219-235	40th to <60th percentile
236-255	60th to <80th percentile
>=256	80th percentile and above

Table 24*Grade 4 Winter*

Scale Score	Percentile
<=156	1
157-161	2
162-166	3
167-171	4
172-175	5
176-177	6
178-180	7
181-182	8
183-184	9
185-186	10
187-188	11
189	12
190-191	13
192	14
193-194	15
195	16
196-197	17
198	18
199-200	19
201	20
202	21
203	22
204	23
205	24
206	25
207	26
208	27
209	28
210	29
211	30
212	31
213	32
214	33
215	35
216	36
217	37
218	38
219	39
220	40

221-238	40th to <60th percentile
239-257	60th to <80th percentile
>=258	80th percentile and above

Table 25*Grade 4 Spring*

231-257	40th to <60th percentile
>=258	60th percentile and above

Scale Score	Percentile
<=161	1
162-167	2
168-172	3
173-177	4
178-181	5
182-183	6
184-186	7
187-188	8
189-190	9
191-192	10
193-194	11
195	12
196-197	13
198	14
199-200	15
201	16
202-203	17
204	18
205-206	19
207	20
208-209	21
210	22
211	23
212	24
213	25
214	26
215	27
216	28
217	29
218	30
219-220	31
221	32
222	33
223	34
224-225	35
226	36
227	37
228	38
229-230	39

Table 26*Grade 5 Fall*

Scale Score	Percentile
<=163	1
164-168	2
169-174	3
175-180	4
181-184	5
185-186	6
187-188	7
189-190	8
191-192	9
193-194	10
195-196	11
197	12
198-199	13
200-201	14
202	15
203-204	16
205	17
206-207	18
208	19
209-210	20
211	21
212	22
213	23
214	24
215	25
216	26
217	27
218	28
219	29
220	30
221-222	31
223	32
224	33
225-226	34
227	35
228	36
229	37
230-231	38
232	39

233	40
234-257	40th to <60th percentile
>=258	60th percentile and above

Table 27*Grade 5 Winter*

237-257	40th to <60th percentile
>=258	60th percentile and above

Scale Score	Percentile
<=165	1
166-171	2
172-177	3
178-184	4
185-188	5
189-190	6
191-192	7
193	8
194-195	9
196-197	10
198-199	11
200	12
201-202	13
203-204	14
205	15
206-207	16
208-209	17
210	18
211-212	19
213	20
214	21
215	22
216	23
217	24
218	25
219	26
220	27
221	28
222	29
223	30
224-225	31
226	32
227	33
228-229	34
230	35
231-232	36
233	37
234	38
235-236	39

Table 28*Grade 5 Spring*

Scale Score	Percentile
<=164	1
165-172	2
173-179	3
180-186	4
187-191	5
192-194	6
195-196	7
197-199	8
200-202	9
203-204	10
205	11
206-207	12
208	13
209-210	14
211	15
212-213	16
214	17
215-216	18
217	19

218	20
219	21
220	22
221	23
222	24
223	25
224-225	26
226	27
227-228	28
229-230	29
231-232	30
233-234	31
235-237	32
238-240	33
241-243	34
244-245	35
246-248	36
249-251	37
252-253	38
254-256	39
257	40
>=258	40th percentile and above

7.0 VALIDITY EVIDENCE: RELATION TO EXTERNAL VARIABLES

To establish the concurrent validity of the Add+Vantage Math Screener, scores were correlated with the i-Ready Diagnostic (Mathematics). The i-Ready assessment serves as an independent, nationally normed criterion measure of general mathematics achievement. The two assessments were created by different vendors, based on different theoretical frameworks of mathematics learning, and piloted on distinct samples. To our knowledge, they have no item overlap.

7.1 Data Linkage and Analytic Sample

Student records from spring of the 2022-2023 and 2023-2024 school years were linked using unique student identifiers. The analytic sample was restricted to student records containing both a valid Add+Vantage Math Screener scale score and a valid i-Ready scale score given in the spring, which was the largest overlap in administrations.

7.2 Concurrent Validity Results

Pearson product-moment correlation coefficients were calculated between the Add+Vantage Math Screener composite scale score and the i-Ready Overall Scale Score (see Table 29).

Table 29

Correlations Between Add+Vantage Math Screener and iReady Overall Scale Score by Grade for Spring Administrations

Grade	N	Correlation [95% CI]
1	653	0.81 [0.79, 0.84]
2	595	0.79 [0.76, 0.82]
3	1,189	0.81 [0.79, 0.83]
4	1,123	0.81 [0.79, 0.83]
5	1,115	0.76 [0.74, 0.79]
Overall	4,675	0.87 [0.86, 0.88]

*Note: All correlations are statistically significant at the $p < .001$ level.

8.0 CLASSIFICATION ACCURACY

Classification accuracy was evaluated to determine the Add+Vantage Math Screener’s effectiveness in identifying students at risk for low mathematics achievement.

8.1 Criterion and Risk Construct

“At-risk” status was defined as scoring at or below the 20th national percentile rank on the i-Ready Mathematics assessment. This norm-referenced benchmark identifies students likely in need of intensive intervention.

8.2 Receiver Operating Characteristic (ROC) Analysis

ROC curve analyses were conducted for each grade level using the Add+Vantage Math Screener rounded composite scale score as the predictor. AUC values are reported with 95% confidence intervals using the DeLong method. Discrimination was considered strong when the lower bound of the 95% CI for the AUC was at or above .80.

8.2.1 Cut Score Decision Rule

For each grade, an optimal Add+Vantage Math Screener cut score was identified using the following hierarchical rule:

- Primary rule. Among all candidate cut points, identify those satisfying both Sensitivity $\geq .80$ and Specificity $\geq .80$, and select the cut point that maximizes Sensitivity (with Specificity as a tiebreaker).
- Fallback rule. If no cut point satisfies the primary rule, select the cut point that maximizes Youden’s J (Sensitivity + Specificity – 1).

The decision rule applied for each grade is reported in the table below.

8.3 Analytic Sample

Analyses were restricted to Grades 1–5 in spring, the only grade $\times$ window cells with sufficient matched cases. Kindergarten was excluded due to insufficient matched records. Fall and winter administrations were excluded for the same reason.

8.4 Results: Spring Administration

Across Grades 1–5, the Add+Vantage Math Screener demonstrated strong overall discrimination, with all AUC point estimates between .85 and .91 and all 95% CI lower bounds at or above .83. Cut scores increased monotonically across grades, consistent with Add+Vantage Math Screener’s vertical scaling and developmental growth interpretations (see **Table 30**).

For Grades 1, 3, and 4, a cut point was identified that simultaneously met the sensitivity $\geq .80$ and specificity $\geq .80$ thresholds. For Grades 2 and 5, no cut point in the data simultaneously satisfied both thresholds, despite strong AUC values (.88 and .85, respectively). For these grades, the cut point maximizing Youden’s J was reported, which produced the most balanced operating-point sensitivity and specificity that the data supported.

In this sample, the proportion of students flagged as at-risk based on the iReady percentile criterion ranged from 32% to 36%, somewhat higher than the nominal 20% national base rate.

Because ROC-based discrimination indices are based on rank ordering, they are robust to base-rate variation across populations.

8.5 Interpretation of Cut Scores

Optimal Add+Vantage Math Screener cut scores were selected using a hierarchical decision rule. The primary rule identified the cut score that maximized Sensitivity among candidate cut points satisfying both Sensitivity $\geq .80$ and Specificity $\geq .80$. When no cut point satisfied the primary rule, a fallback rule selected the cut score maximizing Youden's J (Sensitivity + Specificity - 1). The primary rule was applied for Grades 1, 3, and 4, and the fallback rule was applied for Grades 2 and 5.

Cut scores increased monotonically across Grades 1–5, consistent with the vertical scaling of the Add+Vantage Math Screener and supporting developmental growth interpretations across the elementary span.

Table 30*Classification Accuracy by Grade (Spring)*

Grade	N	AUC [95% CI]	Cut Score*	Sensitivity	Specificity	True Positive	False Positive	False Negative	True Negative	Decision Rule
1	653	.90 [.88, .93]	178.5	.82	.80	184	85	41	343	Primary
2	595	.88 [.84, .91]	193.5	.78	.83	148	67	42	338	Fallback
3	1,189	.91 [.89, .93]	214.5	.86	.81	328	152	55	654	Primary
4	1,123	.89 [.87, .91]	230.5	.84	.83	343	121	63	596	Primary
5	1,115	.85 [.83, .88]	253.5	.79	.85	285	110	77	643	Fallback

**Note.* Cut scores were selected using a hierarchical rule. Primary = the cut point maximizing Sensitivity among those satisfying Sensitivity $\geq .80$ and Specificity $\geq .80$. Fallback = the cut point maximizing Youden's J, applied when no cut satisfied both .80 thresholds. AUC 95% CIs are based on the DeLong method. Add+Vantage Math Screener predictor = rounded composite scale score. Risk criterion = i-Ready Mathematics percentile rank ≤ 20 .

9.0 EVALUATION OF BIAS: CONCURRENT DIFFERENTIAL PREDICTION AND CLASSIFICATION FAIRNESS

This section evaluates potential bias in the Add+Vantage Math Screener by testing for concurrent differential prediction. Specifically, we examined whether the relationship between Add+Vantage at-risk classification and i-Ready proficiency differs across student subgroups (race, gender, and English Learner status) within the same administration window.

9.1 Differential Prediction Framework (Logistic Interaction Models)

Differential prediction was evaluated to examine whether the association between Add+Vantage Math Screener risk classification and i-Ready proficiency differed across student subgroups. “At-risk” was defined as scoring at or below the grade-specific Add+Vantage cut score (Table 30). The i-Ready criterion was overall relative placement, dichotomized as Proficient vs Not Proficient. Within each grade (spring), logistic regression models of the form:

Proficient $\sim$ Add+Vantage_AtRisk $\times$ Group

were estimated separately for EL status (English Learner vs Not an English Learner), gender (female vs male), and race (Non-White vs White). Race/ethnicity was collapsed to White vs Non-White due to sample size limitations across categories within grade. The interaction odds ratio (OR) tested whether Add+Vantage risk status was differentially associated with i-Ready proficiency for the focal group relative to the reference group. When sparse outcome cells were present, Firth logistic regression was used. Interaction p-values were adjusted within each subgroup family across grades using the Benjamini–Hochberg procedure.

9.2 Results of the DTF

After Benjamini–Hochberg adjustment, no interaction effects were statistically significant for EL status, gender, or race, indicating no evidence of differential prediction across these subgroups for spring administrations in Grades 1–5 (see Table 31).

Table 31

Differential Prediction (DTF) Interaction Results by Grade and Subgroup Family (Spring, Grades 1–5)

Subgroup Family	Reference	Grade	N	Model	Interaction OR [95% CI]	p (BH)
EL status	N	1	653	Firth	0.70 [0.18, 2.26]	0.70
EL status	N	2	595	Firth	0.42 [0.00, 3.95]	0.70
EL status	N	3	1189	GLM	1.81 [0.64, 5.10]	0.65
EL status	N	4	1123	GLM	0.99 [0.37, 2.64]	0.98
EL status	N	5	1115	GLM	2.18 [0.81, 5.92]	0.62
Gender	M	1	653	GLM	2.23 [0.82, 6.10]	0.29
Gender	M	2	595	Firth	2.18 [0.53, 12.39]	0.37
Gender	M	3	1189	GLM	2.32 [1.03, 5.26]	0.22
Gender	M	4	1123	GLM	1.53 [0.74, 3.18]	0.37
Gender	M	5	1115	GLM	0.74 [0.35, 1.54]	0.42
Race	White	1	653	GLM	0.54 [0.20, 1.50]	0.30
Race	White	2	595	Firth	0.86 [0.19, 3.37]	0.83
Race	White	3	1189	GLM	0.46 [0.20, 1.07]	0.18
Race	White	4	1123	GLM	0.42 [0.19, 0.93]	0.17
Race	White	5	1115	GLM	0.58 [0.26, 1.28]	0.30

Note. Interaction odds ratios come from grade-specific logistic regression models of the form Proficient ~ Add+Vantage_AtRisk $\times$ Group, where Proficient is i-Ready overall relative placement (Proficient vs Not Proficient) and Add+Vantage_AtRisk is defined as Add+Vantage Math Screener scale score at or below the grade-specific cut score. Firth logistic regression was used when sparse cells were present. Benjamini–Hochberg (BH) adjusted p-values are corrected within each subgroup family across grades.

9.3 Interpretation

Across Grades 1–5, the interaction odds ratios reported in Table 31 test whether the relationship between Add+Vantage Math Screener risk classification and i-Ready proficiency differs across student subgroups. After Benjamini–Hochberg adjustment within each subgroup family, no interaction effects were statistically significant for English Learner (EL) status, gender, or race. Collectively, these results provide no statistically significant evidence of differential prediction by these subgroups.

9.3.1 EL Status

Interaction odds ratios for EL status (EL vs Not an EL) ranged from 0.42 to 2.18 across Grades 1–5, with all 95% confidence intervals including 1.0 and all BH-adjusted p-values $\geq .62$. These

results indicate no statistically significant evidence that the relationship between Add+Vantage risk status and i-Ready proficiency differs between EL and non-EL students.

9.3.2 Gender

Interaction odds ratios for gender (female vs male) ranged from 0.74 to 2.32 across Grades 1–5. Although the unadjusted estimate for Grade 3 was nominally elevated, all BH-adjusted p-values were non-significant ($\geq .22$), and all 95% confidence intervals included or were close to 1.0. These results indicate no statistically significant evidence of differential prediction by gender.

9.3.3 Race

Interaction odds ratios for race (Non-White vs White) ranged from 0.42 to 0.86 across Grades 1–5, with all 95% confidence intervals including 1.0 and all BH-adjusted p-values $\geq .17$. Although several point estimates were below 1.0, the wide confidence intervals indicate substantial uncertainty in the magnitude of any subgroup difference, and the data are consistent with no meaningful difference in predictive relationship between Non-White and White students.

9.4 Overall

Taken together, the differential prediction analyses provide no statistically significant evidence that Add+Vantage Math Screener risk classification operates differently for these subgroups in predicting i-Ready proficiency in spring administrations. These findings support the use of the Add+Vantage Math Screener as a fair classification tool across EL status, gender, and racial subgroups in Grades 1–5, within the limits of the present sample.

10.0 REFERENCES

Vygotsky, L. S. (1981). *Mind in society: The development of higher psychological processes*.

Harvard University Press.

Wright, R. J., & Ellemor-Collins, D. (2018). *The learning framework in number: Pedagogical tools for assessment and instruction* (1st ed.). Sage.