Acadience Learning Inc.

Acadience™ Math Technical Adequacy Brief

Suggested Citation

Gray, J. S., Warnock, A. N., Dewey, E. N., Latimer, R., & Wheeler, C. E. (2019) *Acadience™ Math Technical Adequacy Brief.* Eugene, OR: Acadience Learning Inc.

Technical Adequacy Information Update for Acadience Math

Since Acadience Math (Wheeler et. al., 2019) was first developed, there have been a number of research studies meant to examine and establish its technical adequacy. This document will serve as a quick reference for those looking for technical information on Acadience Math. Technical information provided in this document includes reliability, criterion validity, and classification accuracy. For a timeline showing the schedule of administration for Acadience Math, see Appendix 1 at the end of this document.

Descriptions of the Samples

Sample 1. Validity data for kindergarten and first grade and reliability data for kindergarten Early Numeracy, first grade Early Numeracy, and first, second, fourth, and fifth grade Computation were collected during the 2012–2013 school year. Students in kindergarten and first grade completed both validity and reliability assessment. For validity, the criterion of interest was the Group Mathematics Assessment and Diagnostic Evaluation (GMADE; Williams, 2004). There were 987 students across four schools in four districts in four US states in this sample. Demographic information is not available for this sample.

Sample 2. Reliability data for third and sixth grade Computation and second through sixth grade Concepts and Applications were collected during the 2014–2015, 2015–2016, 2016–2017, and 2018–2019 school years. There were 1,810 students in second through sixth grade across 17 schools in 14 districts in 10 US states in this sample. Demographic information is not available for this sample.

Sample 3. Additional validity data were collected during the 2017–18 school for second through sixth grade. The criterion in this study was the Stanford Achievement Test Series, Tenth

Edition–Total Math score (SAT10; Pearson, 2003). This sample included 537 students across five schools in four districts in four US states. Demographic information is not available for this sample.

Analyses

Reliability. Three types of reliability are reported for Acadience Math: inter-rater, test-retest, and alternate-form. Inter-rater reliability indicates the extent to which results for a measure generalize across assessors. Two methods were used for evaluating inter-rater reliability. For the Early Numeracy measures (i.e., Beginning Quantity Discrimination, Number Identification Fluency, Next Number Fluency, Advanced Quantity Discrimination, and Missing Number Fluency), randomly selected students were administered the measures and were scored simultaneously by two assessors. The two scores were then correlated. For the Computation and Concepts and Applications measures, photocopies were made of unscored student worksheets. The two copies (original and photocopy) were then scored separately and independently by two Acadience Learning research assistants, and the two scores were correlated. Test-retest reliability is an index of score stability when the same test form is administered twice within a short interval of time. Students were administered the same test form twice within an approximate two-week time period and the two scores were correlated. Alternate-form reliability indicates the extent to which test results generalize to different item samples. Students were tested with two different (i.e., alternate), but equivalent, forms of the same measure within a two-week time period and the scores were correlated. Reliability for the inter-rater, test-retest, and alternate-form reliability is reported in Table 1.

Criterion-Related Validity. Concurrent and predictive criterion-related validity of Acadience Math is presented as the correlation between an Acadience Math measure and the criterion measure. The GMADE (Williams, 2004) was used as the criterion for kindergarten and first grade, and the SAT10 Total Math Score (Pearson, 2003) was used as the criterion for second through sixth grade. Because the criteria measures were administered at the end of the school year, concurrent validity is the correlation between the end-of-year Acadience Math measures and the criterion. Predictive validity is the correlation between the Acadience Math measures earlier in the year and the criterion. Both the correlations between the beginning-of-year and middle-of-year Acadience Math scores and the end-of-year criterion score could be considered predictive, so the larger of the two was chosen to report. Criterion-related validity coefficients are presented in Table 2.

Classification Accuracy. Two target outcomes were chosen to evaluate the classification accuracy of Acadience Math measures. The first outcome evaluated Acadience Math with respect to the ability to predict which students would be in need of intensive intervention at the end of the year (i.e., below the 20th percentile on the criterion measure). Second, we evaluated the ability of Acadience Math measures to predict students who were making adequate progress at the end of the year (i.e., above the 40th percentile on the criterion). The end-of-year Acadience Math Measures were used to assess classification accuracy. The GMADE (Williams, 2004) was used as the criterion for kindergarten and first grade, and the SAT10 Total Math Score (Pearson, 2003) was used as the criterion for second through sixth grade. Classification accuracy was assessed using four criteria: sensitivity, specificity, AUC, and Cohen's Kappa. Sensitivity is the proportion of true positives (e.g., students who were in need of

intensive intervention and identified as such). Specificity is the proportion of true negatives (e.g., students who were correctly identified as not needing intensive intervention). The tradeoff between sensitivity and specificity is visualized in the receiver operator characteristic (ROC) curve. The area under this curve (AUC) is another measure of classification that quantifies the extent of this tradeoff. Lastly, Cohen's Kappa is used to quantify the association between two judgments or ratings, in this case the agreement between the Acadience Math measure and the criterion (Cohen, 1960). For all of these statistics, the maximum value is 1.0, with higher values indicating greater classification accuracy. Classification accuracy is presented in Tables 3 and 4.

Table 1. *Inter-Rater, Test-Retest, and Alternate-Form Reliability Estimates for Acadience Math*

	Grade						
Acadience Math Measure	K	1	2	3	4	5	6
Beginning Quantity Discrimination							
Inter-Rater	.99	--	--	--	--	--	--
Test-Retest	.76	--	--	--	--	--	--
Alternate-Form	.63	--	--	--	--	--	--
Number Identification Fluency							
Inter-Rater	.99	.99	--	--	--	--	--
Test-Retest	.93	.94	--	--	--	--	--
Alternate-Form	.88	.90	--	--	--	--	--
Next Number Fluency							
Inter-Rater	.99	.99	--	--	--	--	--
Test-Retest	.76	.81	--	--	--	--	--
Alternate-Form	.80	.66	--	--	--	--	--
Advanced Quantity Discrimination							
Inter-Rater	--	.99	--	--	--	--	--
Test-Retest	--	.86	--	--	--	--	--
Alternate-Form	--	.88	--	--	--	--	--
Missing Number Fluency							
Inter-Rater	--	.99	--	--	--	--	--
Test-Retest	--	.87	--	--	--	--	--
Alternate-Form	--	.82	--	--	--	--	--
Computation[1]							
Inter-Rater	--	.99	.99	.98	.99	.99	.99
Two-Week Test-Retest	--	.73	.66	.81	.87	.76	.80
Total Score Test-Retest	--	.81	.77	.86	.90	.81	.84
Two-Worksheet Alternate-Form	--	.79	.75	.86	.89	.75	.83
Two-Week Alternate-Form	--	.59	.69	.82	.83	.60	.74
Total Score Alternate-Form	--	.67	.78	.88	.87	.73	.81
Concepts and Applications							
Inter-Rater	--	--	.99	1.00	.99	.99	.99
Test-Retest	--	--	.75	.75	.85	.75	.72
Alternate-Form	--	--	.74	.79	.87	.78	.88

Note. Based on Samples 1 and 2. Dashes indicate the measure is not given at the specified time of year to the grade level.

[1] At each Computation testing session, students completed two worksheets (i.e., Worksheet A and Worksheet B or Worksheet C and D). For test-retest reliability, students completed the same worksheets at time one and time two (i.e., Worksheets A and B at time one, Worksheets

A and B again at time two). For alternate-form reliability, students completed an alternate, but equivalent, set of worksheets during the second testing session (i.e., Worksheet A and Worksheet B at time one, Worksheet C and Worksheet D at time two). For each testing session, scores were calculated for the two worksheets (e.g., Worksheet A Total Score) and the two scores were then averaged to calculate a Total Score. Two-week test-retest reliability for Computation is the median of the two possible pairs of test-retest worksheets (i.e., Worksheet A with Worksheet A and Worksheet B with Worksheet B). Total Score test-retest reliability for Computation is the correlation between the Total Score from the first testing session and the Total Score from the second testing session. Two-worksheet alternate-form reliability for Computation is the median of the two possible pairs of worksheets administered at the same time point (i.e., Worksheet A with Worksheet B and Worksheet C with Worksheet Form D). Two-week alternate-form reliability for Computation is the median of four possible pairs of two-week alternate worksheets (i.e., Worksheet A with Worksheet C, Worksheet A with Worksheet D, Worksheet B with Worksheet C, and Worksheet B with Worksheet D). Total Score alternate-form reliability for Computation is the correlation between the Total Score from the first testing session and the Total Score from the second testing session.

Table 2. *Predictive and Concurrent Criterion-Related Validity for Acadience Math*

Acadience Math Measure	Grade						
	K	1	2	3	4	5	6
Beginning Quantity Discrimination							
Predictive	.39	--	--	--	--	--	--
Concurrent	.45	--	--	--	--	--	--
Number Identification Fluency							
Predictive	.29	.59	--	--	--	--	--
Concurrent	.38	NA	--	--	--	--	--
Next Number Fluency							
Predictive	.36	.55	--	--	--	--	--
Concurrent	.31	NA	--	--	--	--	--
Advanced Quantity Discrimination							
Predictive	--	.64	--	--	--	--	--
Concurrent	--	.65	--	--	--	--	--
Missing Number Fluency							
Predictive	--	.58	--	--	--	--	--
Concurrent	--	.55	--	--	--	--	--
Computation							
Predictive	--	.45	.72	.72	.71	.73	.84
Concurrent	--	.55	.68	.69	.72	.76	.82
Concepts and Applications							
Predictive	--	--	.79	.74	.81	.81	.87
Concurrent	--	--	.84	.83	.76	.81	.84
Math Composite Score							
Predictive	.40	.63	.80	.81	.81	.81	.86
Concurrent	.46	.65	.83	.81	.77	.83	.87

Note. Based on Samples 1 and 3. Dashes indicate the measure is not given at the specified time of year to the grade level. The GMADE was used as the criterion for kindergarten and first grade, and the SAT10 Total Math Score was used as the criterion for second through sixth grade. Concurrent validity is the correlation between the end-of-year Acadience Math measure and the criterion (administered at end of year). Predictive validity is the larger of the two correlations between the beginning-of-year or middle-of-year Acadience Math and the criterion (administered at end of year). Concurrent validity is not available for Number Identification Fluency or Next Number Fluency in first grade because those measures are not administered at the end of the year.

Table 3. *Classification Accuracy of Acadience Math for Intensive Support Outcome*

Acadience Math Measure	Grade						
	K	1	2	3	4	5	6
Beginning Quantity Discrimination							
Sensitivity	.58	--	--	--	--	--	--
Specificity	.84	--	--	--	--	--	--
AUC	.78	--	--	--	--	--	--
Kappa	.43	--	--	--	--	--	--
Number Identification Fluency							
Sensitivity	.58	.49	--	--	--	--	--
Specificity	.88	.88	--	--	--	--	--
AUC	.72	.80	--	--	--	--	--
Kappa	.45	.38	--	--	--	--	--
Next Number Fluency							
Sensitivity	.47	.50	--	--	--	--	--
Specificity	.96	.89	--	--	--	--	--
AUC	.75	.83	--	--	--	--	--
Kappa	.45	.41	--	--	--	--	--
Advanced Quantity Discrimination							
Sensitivity	--	.53	--	--	--	--	--
Specificity	--	.93	--	--	--	--	--
AUC	--	.85	--	--	--	--	--
Kappa	--	.47	--	--	--	--	--
Missing Number Fluency							
Sensitivity	--	.52	--	--	--	--	--
Specificity	--	.82	--	--	--	--	--
AUC	--	.77	--	--	--	--	--
Kappa	--	.35	--	--	--	--	--
Computation							
Sensitivity	--	.54	.43	.46	.50	.46	.44
Specificity	--	.92	.95	.86	.95	.97	.98
AUC	--	.80	.80	.76	.78	.78	.81
Kappa	--	.49	.43	.33	.49	.47	.47
Concepts and Applications							
Sensitivity	--	--	.46	.50	.46	.49	.41
Specificity	--	--	.99	.95	.95	.97	.99

Acadience Math Measure	Grade						
	K	1	2	3	4	5	6
AUC	--	--	.85	.86	.78	.83	.91
Kappa	--	--	.55	.50	.47	.50	.47
Math Composite Score							
Sensitivity	.52	.51	.43	.49	.48	.51	.42
Specificity	.98	.88	.97	.91	.95	.99	.99
AUC	.72	.87	.84	.82	.78	.83	.87
Kappa	.49	.41	.47	.42	.49	.55	.49

Note. Based on Samples 1 and 3. Dashes indicate the measure is not given at the specified time of year to the grade level. Intensive Support Outcome indicates students who were below the 20th percentile on the criterion measure at the end of year. The GMADE was used as the criterion for kindergarten and first grade, and the SAT10 Total Math Score was used as the criterion for second through sixth grade. End-of-year Acadience Math scores were used to evaluate classification accuracy.

Table 4. *Classification Accuracy of Acadience Math for Adequate Progress Outcome*

	Grade						
Acadience Math Measure	K	1	2	3	4	5	6
Beginning Quantity Discrimination							
Sensitivity	.91	--	--	--	--	--	--
Specificity	.52	--	--	--	--	--	--
AUC	.76	--	--	--	--	--	--
Kappa	.20	--	--	--	--	--	--
Number Identification Fluency							
Sensitivity	.78	.89	--	--	--	--	--
Specificity	.56	.67	--	--	--	--	--
AUC	.69	.74	--	--	--	--	--
Kappa	.18	.36	--	--	--	--	--
Next Number Fluency							
Sensitivity	.74	.93	--	--	--	--	--
Specificity	.68	.57	--	--	--	--	--
AUC	.72	.77	--	--	--	--	--
Kappa	.26	.29	--	--	--	--	--
Advanced Quantity Discrimination							
Sensitivity	--	.82	--	--	--	--	--
Specificity	--	.68	--	--	--	--	--
AUC	--	.75	--	--	--	--	--
Kappa	--	.34	--	--	--	--	--
Missing Number Fluency							
Sensitivity	--	.96	--	--	--	--	--
Specificity	--	.52	--	--	--	--	--
AUC	--	.78	--	--	--	--	--
Kappa	--	.26	--	--	--	--	--
Computation							
Sensitivity	--	.93	.87	.91	.75	.82	.82
Specificity	--	.45	.72	.66	.83	.75	.80
AUC	--	.78	.86	.81	.81	.85	.87
Kappa	--	.20	.41	.46	.52	.50	.53
Concepts and Applications							
Sensitivity	--	--	.91	.90	.94	.82	.94
Specificity	--	--	.88	.70	.74	.74	.80

	Grade						
Acadience Math Measure	K	1	2	3	4	5	6
AUC	--	--	.93	.88	.91	.84	.93
Kappa	--	--	.67	.49	.52	.49	.61
Math Composite Score							
Sensitivity	.61	.99	.91	.90	.88	.85	.88
Specificity	.75	.58	.88	.73	.79	.82	.81
AUC	.72	.80	.92	.85	.89	.87	.91
Kappa	.25	.33	.67	.53	.56	.61	.60

Note. Based on Samples 1 and 3. Dashes indicate the measure is not given at the specified time of year to the grade level. Adequate Progress Outcome indicates students who were above the 40th percentile on the criterion measure at the end of year. The GMADE was used as the criterion for kindergarten and first grade, and the SAT10 Total Math Score was used as the criterion for second through sixth grade. End-of-year Acadience Math scores were used to evaluate classification accuracy.

Appendix 1. Timeline of Administration for Acadience Math Measures

Acadience Math Measure and Time of Year	Grade K	1	2	3	4	5	6
Beginning Quantity Discrimination							
Beginning of Year	X	--	--	--	--	--	--
Middle of Year	X	--	--	--	--	--	--
End of Year	X	--	--	--	--	--	--
Number Identification Fluency							
Beginning of Year	X	X	--	--	--	--	--
Middle of Year	X	--	--	--	--	--	--
End of Year	X	--	--	--	--	--	--
Next Number Fluency							
Beginning of Year	X	X	--	--	--	--	--
Middle of Year	X	--	--	--	--	--	--
End of Year	X	--	--	--	--	--	--
Advanced Quantity Discrimination							
Beginning of Year	--	X	--	--	--	--	--
Middle of Year	--	X	--	--	--	--	--
End of Year	--	X	--	--	--	--	--
Missing Number Fluency							
Beginning of Year	--	X	--	--	--	--	--
Middle of Year	--	X	--	--	--	--	--
End of Year	--	X	--	--	--	--	--
Computation							
Beginning of Year	--	X	X	X	X	X	X
Middle of Year	--	X	X	X	X	X	X
End of Year	--	X	X	X	X	X	X
Concepts and Applications							
Beginning of Year	--	--	X	X	X	X	X
Middle of Year	--	--	X	X	X	X	X
End of Year	--	--	X	X	X	X	X

Note. Dashes indicate the measure is not given at the specified time of year to the grade level.

References

Acadience Learning (2019). *Acadience Math Benchmark Goals and Composite Score*. Available: www.acadiencelearning.org.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Pearson. (2003). Stanford Achievement Test Series, Tenth Edition (SAT10). San Antonio, TX.

Wheeler, C. E., Lembke, E., S., Richards-Tutor, C., Wallin, J., Good, R. H., III, Dewey, E. N., & Warnock, A. N. (2019). Acadience Math. Eugene, OR: Acadience Learning.

Williams, K. T. (2004). Group Math Assessment and Diagnostic Evaluation Assessment (GMADE). New York: Pearson.