Technical Report for Pilot Study of Content Area Multisyllabic Word Identification Formative Assessments: Summary Results

Yaacov Petscher, PhD and Sarah Kershaw, PhD

This summary presents the findings of a pilot study which examined the preliminary reliability and convergent validity of the three multisyllabic word identification assessment measures that make up the Word ID Formative Assessment System. The pilot study had two aims: 1) to evaluate the reliability (or internal consistency) of assessment items and 2) to evaluate the convergent validity of total scores. For the pilot study, 112 students in Grades 6, 8, 9, and 10 were randomly selected and assigned to one of four content area assessments: English Language Arts (ELA), Math, Science, or Social Studies. This resulted in a sample size of 28 in each content area.

RELIABILITY RESULTS

Reliability refers to the consistency with which an instrument measures a student's ability. A reliability coefficient for any given set of scores from an assessment must approximate or exceed .80 to be considered acceptable for the purpose of research; coefficients of .90 or higher are considered the most desirable for clinical decision making (Aiken and Groth-Marnat 2006; Nunnally and Bernstein 1994; Salvia, Ysseldyke, and Bolt 2007). Different types of reliability coefficients may be computed. The results summarized herein use coefficient alpha, computed using Cronbach's (1951) method, as the measure of internal consistency for all assessment tasks.

Maze Measure   As shown in Table 1, Cronbach's alphas for the Maze Measure tasks exceed the coefficient minimum of .80 for all content areas (α = .91 for ELA, α = .86 for Math, α = .94 for Science, and α = .84 for Social Studies).

20/20 Screener   Table 2 reports the reliability estimates by content area for each list of 20 names and 20 terms on the 20/20 Screener. As shown in this table, estimated reliabilities indicate that each list of 20 items measuring correct phonetic pronunciation exceed the minimum criterion for reliability (i.e., α = .80).

360° Diagnostic Exam   Table 3 reports the reliability estimates by content area for the names and terms tasks on the 360° Diagnostic Exam. As shown in this table, estimated reliabilities indicate that the sets of items measuring correct phonetic pronunciation exceed the minimum criterion for reliability (i.e., α = .80).

VALIDITY RESULTS

The validity of an assessment tool refers to its accuracy, or the extent to which a test measures the construct it purports to measure. Scores from the Maze Measure and from the 20/20 Screener names and terms tasks were compared with those of the names and terms tasks on the 360° Diagnostic Exam in order to determine the convergent validity among the three assessment measures. The goal in comparing these scores was to ensure that all three measures were working together to provide an accurate account of students' performance.

Table 4 reports the correlations within each content area. Within ELA, the correlations were very strong among all three tasks; correlations ranged from .60 between the 20/20 terms task and the 360° names task to .81 between the Maze Measure and the 360° terms task. This suggests that performance on the Maze Measure screener is associated with performance on the names and terms tasks on both the 20/20 Screener

Table 1. Reliability Estimates for the Maze Measure by Content Area

MAZE MEASURE

Content Area	Number of items	α
ELA	33	0.91
Math	40	0.86
Science	37	0.94
Social Studies	38	0.84

Note. ELA = English Language Arts.
α = Cronbach's alpha reliability coefficient.

Table 2. Reliability Estimates for the 20/20 Screener by Content Area

20/20 SCREENER

Content Area	Names α	Terms α
ELA	0.90	0.90
Math	0.87	0.88
Science	0.98	0.99
Social Studies	0.93	0.94

Note. ELA = English Language Arts.
α = Cronbach's alpha reliability coefficient.

Table 3. Reliability Estimates for the 360° Diagnostic Exam by Content Area

360° DIAGNOSTIC EXAM

Content Area	Names # of items	α	Terms # of items	α
ELA	38	0.90	56	0.91
Math	36	0.88	56	0.92
Science	50	0.98	58	0.99
Social Studies	42	0.93	56	0.95

Note. ELA = English Language Arts.
α = Cronbach's alpha reliability coefficient.

Table 4. Correlations Among the Maze Measure, 20/20 Screener, and the 360° Diagnostic Exam by Content Area

ELA

Task	Maze	360° Names	360° Terms
Maze	1.00	.65***	.81***
20/20 Names	.61***	1.00	.72***
20/20 Terms	.81***	.60***	1.00

MATH

Task	Maze	360° Names	360° Terms
Maze	1.00	.55**	.52**
20/20 Names	.43*	1.00	.86***
20/20 Terms	.53**	.82***	1.00

SCIENCE

Task	Maze	360° Names	360° Terms
Maze	1.00	.85***	.84***
20/20 Names	.79***	1.00	.96***
20/20 Terms	.82***	.98***	1.00

SOCIAL STUDIES

Task	Maze	360° Names	360° Terms
Maze	1.00	.60***	.63***
20/20 Names	.62***	1.00	.94***
20/20 Terms	.61***	.92***	1.00

Note. ELA = English Language Arts. *p < .05; **p < .01; ***p < .001.

and the 360° Diagnostic Exam. Similarly, moderate to strong correlations were estimated among all three tasks for Math (range of .43 to .86), Science (range of .79 to .98), and Social Studies (range of .60 to .94).

CONCLUSION

In this pilot study, a high degree of reliability was achieved for the Maze Measure, the 20/20 Screener, and the 360° Diagnostic Exam—all three assessments appear to be internally consistent. Convergent validity analyses also yielded moderate to strong coefficients for the sample in this pilot study. It should be noted that these results are preliminary; future studies should be carried out with a larger sample in order to replicate the estimates found here and also to evaluate other assessment factors.