Technical Information

History

The current passages were originally developed in two phases. In each phase, a different type of reliability and validity study was conducted. The first phase resulted in the Reading Fluency Monitor (RFM) first edition, and the second phase resulted in the RFM second edition (now called the Reading Fluency Benchmark Assessor).

To create the first edition, the benchmark passages were developed to match the readability of anchor passages that had been extensively field-tested. The data for single-measure technical adequacy were collected primarily in Washington State from 1998 to 2000.* The data and correlations from the initial anchor passages provided substantial evidence of both concurrent and predictive validity.

In order to further enhance the RFM benchmark passages, Read Naturally conducted extensive field-testing in which students read the passages orally for one-minute timings. This testing ensured that the passages within each grade level are similar in difficulty. The data collected during the field testing indicated that the scores of the new passages resulted in a narrower range of scores at the intended grade level across sets of three passages. Therefore, these new passages replaced a number of the first-edition benchmark passages. Multiple-measure reliability and validity studies were conducted during the 2002–2003 school year in Minnesota, California, Texas, Virginia, Michigan, Iowa, and Pennsylvania.

The following information provides evidence for the reliability and validity of the procedures employed in the current version of the Reading Fluency Benchmark Assessor. All data were collected in public school classrooms. The reliability and validity coefficients reported in this section are based on words read correctly per minute on standard passages.

*The technical information from the RFM first edition is available on the Read Naturally website.

Reliability

Reliability is the degree to which scores produced under standard procedures will be dependably replicated on another occasion, with sets of similar passages, and with different test administrators. Evidence for the single measure reliability of oral reading fluency measurement is provided in Table 1 and Table 2. The 121 correlations contributing to the database were meta-analyzed using the Comprehensive Meta-Analysis software (Borenstein and Rothstein, 1999). The results of this analysis are presented in two ways. Table 1 represents test-retest reliability coefficients for groups of children in grades one through seven leveled by three time-span categories as follows:

- **Reliability:** Test and re-test were done within the same academic year;
- **Reliability-Delayed:** Test and re-test were done in contiguous years;
- **Reliability-Delayed 2:** Test and re-test were separated by more than one year.

The data in Table 1 show that across the 121 coefficients, the point estimate is .907 with a 95% confidence interval from .899 to .915. This is well within acceptable psychometric standards. The magnitude of the correlations is remarkably stable across time-span categories, although it is not surprising to note a slight inverse relationship between the size of the correlation and the length of time between test administrations.

Table 1: Meta-analysis of Oral Reading Fluency Reliability Coefficients

Time: Same Year or Delayed	Effect	Lower	Upper	N Total	P Value
Same Year (84)*	.915	.906	.923	12213	.000
Delayed: Consecutive Years (19)*	.896	.881	.909	958	.000
Delayed 2: Alternate Years (18)*	.866	.820	.901	1361	.000
Total Combined (121)*	.907	.899	.915	14532	.000

* ( ) number of correlation coefficients in meta-analysis

Table 2 shows the single measure reliability coefficients organized by grade level across time span categories. The grade one reliability estimate (.847) is lower than that for the other grade levels and can be explained by the fact that many first-grade children are just beginning to learn to read. The reliability estimates for all other grades are above .90.

Table 2: Meta-analysis of Oral Reading Fluency Reliability Coefficients by Grade Level

Grade	Effect	Lower	Upper	N Total	P Value
1 (8)*	.847	.768	.901	1631	.000
2 (16)*	.914	.895	.929	3920	.000
3 (23)*	.909	.897	.920	4345	.000
4 (38)*	.905	.883	.922	2121	.000
5 (28)*	.916	.905	.925	2131	.000
6 (2)*	.922	.877	.946	122	.000
7 (6)*	.906	.861	.936	262	.000
Combined (121)*	.907	.899	.915	14532	.000

* ( ) number of correlation coefficients in meta-analysis

The data in Table 3 show reliability coefficients for three oral reading passages per grade in the fall, winter, and spring of the year. After each round of data collection, items that were discrepant in difficulty (i.e., too easy or too hard) were modified. Thus, the intraclass correlation, which takes into account the difficulty of passages, increased over time. Two types of coefficients are presented: (a) Cronbach's Alpha, a measure of internal consistency among the three passages and (b) An Intra-class (Brennan, 2000) correlation, an index that takes into account absolute differences in performance. In both cases, the correlation coefficient can range from 0 (i.e., no correlation) to 1.0 (i.e., a perfect correlation). The appropriate confidence intervals ("CI") are reported in the columns next to each coefficient. All correlations were statistically significant at the .0001 level.

The Standard Error of Measurement

Even with substantial reliability, one must be sure to recognize that there is error involved in any single score. The magnitude of this error is quantified with a value called the standard error of measurement. The size of the standard error of measurement is a function of the standard deviation of the measure and the reliability estimate.

In the Reading Fluency Benchmark Assessor research with single passages, the standard deviation of oral reading fluency scores was in the neighborhood of 35 to 42. Using a reliability coefficient of .90, the standard error of measurement was estimated to be somewhere between 11 and 13.

Table 4 shows the obtained means, standard deviations and standard errors for the three passages standardized for the spring 2002–2003 samples. The mean of three virtually equivalent passages has considerably less measurement error than the single passage.

Table 3: Reliability of Three Reading Fluency Benchmark Assessor Passages, Fall, Winter, and Spring 2002–2003 School Year

Grade	Intra-class Correlation*, Single Passage	95% CI	Intra-class Correlation*, 3-Passage Average	95% CI	Number of Students
1 – Fall	.96	.94 to .98	.99	.98 to .99	73
1 – Winter	.96	.93 to .97	.99	.98 to .99	84
1 – Spring	.95	.91 to .97	.98	.97 to .99	84
2 – Fall	.85	.63 to .93	.95	.83 to .97	80
2 – Winter	.96	.95 to .97	.99	.98 to .99	95
2 – Spring	.95	.93 to .97	.98	.98 to .99	90
3 – Fall	.94	.92 to .96	.98	.97 to .99	105
3 – Winter	.93	.91 to .95	.98	.97 to .98	123
3 – Spring	.94	.91 to .95	.98	.97 to .98	121
4 – Fall	.94	.92 to .96	.98	.97 to .98	110
4 – Winter	.92	.89 to .94	.97	.96 to .98	138
4 – Spring	.93	.90 to .95	.97	.97 to .98	120
5 – Fall	.94	.92 to .95	.98	.97 to .98	144
5 – Winter	.91	.89 to .94	.97	.96 to .98	139
5 – Spring	.92	.89 to .94	.97	.96 to .98	158
6 – Fall	.85	.72 to .91	.95	.89 to .97	101
6 – Winter	.90	.86 to .92	.96	.95 to .97	138
6 – Spring	.90	.84 to .93	.96	.94 to .98	86
7 – Fall	.92	.87 to .95	.97	.95 to .98	87
7 – Winter	.92	.89 to .95	.97	.96 to .98	82
7 – Spring	.90	.79 to .95	.96	.92 to .98	84
8 – Fall	.88	.84 to .92	.96	.94 to .97	87
8 – Winter	.92	.89 to .95	.97	.96 to .98	86
8 – Spring	.91	.88 to .94	.97	.96 to .98	115

*The intra-class correlation is calculated from a mixed model Generalizability Analysis (G Study) where students are a random factor and passages are a fixed factor. The model includes variance components for absolute differences among the passages as a source of measurement error. The single measure estimate is from the D study where decisions are made based on information from one of the three Reading Fluency Benchmark Assessor measures.

Table 4: Mean, Standard Deviation and Standard Error of Measurement of Three Reading Fluency Benchmark Assessor Passages, Spring 2002-2003 School Year

Grade	Three-Passage Mean	Standard Deviation	Three-Passage Reliability	Standard Error of Measurement
1	82.9	40.5	.98	8.1
2	121.7	47.1	.98	9.4
3	125.8	45.5	.98	9.1
4	132.8	39.0	.97	9.5
5	142.5	39.7	.98	7.9
6	141.5	34.0	.96	9.5
7	148.3	38.5	.96	10.8
8	150.9	41.3	.97	10.0

Validity and Reading Comprehension

Validity is the degree to which an assessment measures what it purports to measure. In this case, the Reading Fluency Benchmark Assessor purports to measure reading achievement.

Tables 5 and 6 provide initial evidence (1998 to 2000) for the validity of oral reading fluency measures as an indicator of reading comprehension. The combined effects between the oral reading fluency measure and various high-stakes measures of reading were analyzed using the meta-analysis computer program referenced above (Borenstein and Rothstein, 1999). Time between the administration of the oral reading fluency measure and administration of the criterion measure ranged from one month to two years. The analysis was done in two ways: (1) across grades 1 through 6 by criterion measure and (2) across criterion measure by grade level (see Tables 5 and 6, respectively).

The overall validity estimate across grades and measures was .730 with a 95% confidence interval from .716 to .744. The variance interpretation of this correlation is that roughly half of the variability in high-stakes reading comprehension scores can be explained by oral reading fluency.

Table 5: Meta-analysis of Validity Coefficients Using Oral Reading Fluency with Measures of Reading Comprehension

Test	Effect	Lower	Upper	N Total	P Value
CTBS[a] Comprehension (4)[b]	.789	.726	.839	186	.000
CTBS[a] Total Reading (8)[b]	.770	.730	.804	486	.000
CTBS[a] Vocabulary (1)[b]	.855	.745	.920	42	.000
Gates[c] Comprehension (8)[b]	.746	.703	.783	784	.000
Gates[c] Total Reading (36)[b]	.782	.763	.800	2382	.000
Gates[c] Vocabulary (9)[b]	.730	.684	.770	585	.000
ITBS[d] Comprehension (11)[b]	.688	.655	.717	6346	.000
ITBS[d] Total Reading (8)[b]	.712	.677	.743	3297	.000
ITBS[d] Vocabulary (11)[b]	.610	.594	.625	6409	.000
WASL[e] Reading (19)[b]	.680	.647	.711	1136	.000
Combined (115)[b]	.730	.716	.744	21653	.000

[a]CTBS: Comprehensive Test of Basic Skills
[b]( ) number of correlation coefficients in meta-analysis
[c]Gates: Gates MacGinitie Reading Tests (GMRT), 3rd edition
[d]ITBS: Iowa Test of Basic Skills
[e]WASL: Washington Assessment of Student Learning (state-mandated performance assessment)

Table 6: Meta-analysis of Validity Coefficients Using Oral Reading Fluency with Measures of Reading Comprehension by Grade Level

Grade	Effect	Lower	Upper	N Total	P Value
1 (3)*	.779	.716	.829	200	.000
2 (26)*	.731	.702	.758	10626	.000
3 (29)*	.698	.671	.723	7529	.000
4 (29)*	.748	.719	.774	1734	.000
5 (19)*	.752	.724	.778	1075	.000
6 (1)*	.760	.605	.860	47	.000
Combined (115)*	.730	.716	.744	21653	.000

* ( ) number of correlation coefficients in meta-analysis

Recent validity studies (2002–2003) within grade level, found moderate to high predictive validity for the mean of three passages. Time between the administration of the oral reading fluency measure and administration of the criterion measure ranged from one month to one year (see Table 7).

Table 7: Predictive Validity of Three Reading Fluency Benchmark Assessor Passages, Spring 2002–2003 School Year

Test	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Median
MCA[a]		.838 (n=24)		.753 (n=37)				.796
NALT[b]	.883 (n=18)	.927 (n=23)	.911 (n=23)	.787 (n=34)	.766 (n=32)	.820 (n=35)		.852
Stanford-9[c] 1-year lag		.640 (n=13)	.754 (n=23)	.553 (n=25)	.787 (n=26)	.804 (n=29)		.754
ITBS[d] 1-year lag						.649 (n=25)	.321 (n=26)	.485
SDRT[e]						.832 (n=25)		.832
TAAS[f] 1-year lag				.543 (n=23)	.562 (n=19)			.553
PSSA[g]					.806 (n=24)			.806
SOL[h]			.891 (n=22)					.891
Grand Median								.766

[a]Minnesota Comprehensive Assessment
[b]Northwest Achievement Levels Tests
[c]Stanford Achievement Tests – Version 9
[d]Iowa Test of Basic Skills
[e]Stanford Diagnostic Reading Test
[f]Texas Assessment of Academic Skills
[g]Pennsylvania System of School Assessment
[h]Virginia Standards of Learning Assessment

When using oral reading fluency scores to predict performance on a subsequent high-stakes measure, it is important to remember that there is error associated with any such prediction and that the error involved in making predictions about individuals is substantial. Predictions involving groups are somewhat more accurate.

Other evidence for the overall validity of oral reading fluency measures is the observation that children identified in various ways for participation in compensatory educational programs (usually for reading problems) score significantly lower on oral reading fluency measures than children not so identified (Davidson, Blake, & Towner, 1998; Davidson, Myhre, & Towner, 1999; Davidson, M., Stage, S., & Towner, J., 1999; Davidson, M., & Towner, J., 2001; Towner, Davidson, & Howell, 2001). That is, oral reading fluency scores are consistent with more involved identification procedures commonly employed in the schools.

Oral Reading Prediction of Program Placement

Readability Levels of the Passages

The passages used in the Reading Fluency Benchmark Assessor were leveled according to several commonly used readability formulas. Calculations were performed using Readability Calculations (Micro Power and Light, 2000) and Lexile Analysis (Stenner, 2000).

Bibliography

Adams, M. J. (1990). *Beginning to Read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis: A computer program for research synthesis.* Englewood, NJ: Biostat.

Davidson, M., Blake, G., & Towner, J. (1998). *Findings from the second-grade reading testing pilot program.* Bellingham, WA: Applied Research and Development Center, Western Washington University.

Davidson, M., Myhre, O., & Towner, J. (1999). *Victory 1000 final report: Phases I, II, and III.* Bellingham, WA: Applied Research and Development Center, Western Washington University.

Davidson, M., Stage, S., & Towner, J. (1999). *Second grade testing follow-up study.* Bellingham, WA: Applied Research and Development Center, Western Washington University.

Davidson, M., & Towner, J. (2001). The reliability, validity, and applications of oral reading fluency measures. Paper presented at the meeting of the Society of the Scientific Study of Reading, Boulder, CO.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, C., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22* (1), 27–48.

Fuchs, Lynn S., Fuchs, Douglas, Hosp, Michelle K., Jenkins, & Joseph R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5* (3), 239–256.

Hasbrouck, J., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24* (3), 41–44.

Heistad, D. J. (1998). A district-wide, individually administered oral reading assessment in first grade: Technical adequacy and standards-based reporting. Paper presented to American Education Research Association Annual Conference, San Diego, CA.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52* (3), 266–276.

Towner, J., Davidson, M., & Howell, K. (2001). *Evaluating an oral reading fluency cut-score to identify second grade children who are at risk for reading failure.* Bellingham, WA: Applied Research and Development Center, Western Washington University.

Trummer, W. E., Herriman, M. L., & Nesdale A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23*, 134–158.

Walsh, D. J., Price, G. G., & Gillingham, M. G. (1988) The critical, but transitory importance of letter naming. *Reading Research Quarterly, 23*, 108–122.