Standard Reading Assessment Passages (RAPs) For Use in General Outcome Measurement:

A Manual Describing Development and Technical Features

Kathryn B. Howe, Ph.D. & Michelle M. Shinn, Ph.D.

August 2002

6420 Flying Cloud Drive, Suite 204
Eden Prairie, MN 55344
P: (952) 944-1882 (888)944-1882 F: (952) 944-1884
www.edformation.com

Table of Contents

Purpose of the Technical Manual

This manual is intended to provide the user of the ***AIMSweb Standard Reading Assessment Passages (RAPs)*** with background knowledge regarding the development and field testing of the passages and their technical features, including evidence of (1) alternate-form reliability, and (2) equivalent difficulty within any given set of graded passages. Data on the passages are organized at three levels:

1. Technical features of all RAPs within a grade.
2. Technical features of the 3 RAPs used with ***AIMSweb Benchmark;*** and
3. Technical features of the 20-30 RAPs used with ***AIMSweb Progress Monitor.***

The Need for Standard Passages for General Outcome Measurement

Educators have many commercially available reading tests that are intended to describe students' level of reading skills compared to others from a range of comparison groups, and for purposes of accountability. Many of these tests are of high quality when examined from the technical perspective of reliability and validity. In addition, many of these tests do a fine job of what they were principally intended to do, compare the performance of one group of students to the performance of another group of students. For example, the reading portion of the Iowa Test of Basic Skills (ITBS) is well suited to provide a picture of a student's **general** reading skills **at one place** in time compared to students in the school district, in the state, or in the country. When students are sampled appropriately, the ITBS allows school-to-school, school-to-state, and even state-to-state comparisons.

Three items are noteworthy about these kinds of tests. First, it is important to note the approach implicit in most reading testing like the ITBS; this approach is a **static** one, where students are tested once per year. Second, the major purposes of these kinds of tests are for making decisions **other** than those necessary **to inform teaching.** Information from a group-administered, high-stakes reading test will be used to decide if the student moves to the next grade, is retained, or must go to summer school. Information from an individually administered reading test may be used to determine eligibility for a special program. Both group and individual achievement tests are limited in their ability to inform teaching because, too often, they are given after students have had lots of reading instruction (e.g., end of the school year) and too often, the results take a long time to receive. This "after the fact" assessment information is not useful to educators to inform instruction.

Third, these types of reading tests take **a long time to administer** and are quite **expensive.** If 3 half-days are devoted to state testing and 5 half-days are devoted to group achievement testing, almost 1 week of instructional time is lost. One would like to see the loss of instructional time offset by the benefits of receiving the information in a timely manner to inform changes in teaching.

The shortcomings of conventional readings tests are well known and have been remedied by the development of a new type of testing using new types of tests created for the explicit purpose of informing instruction, especially when decisions about students' reading progress is required. This type of reading test is short (1 minute), simple to give and understand, and provides an accurate assessment of students' reading growth.

This reading testing technology is identified in the professional research literature as *Curriculum-Based Measurement CBM (Deno, 1985: Shinn, 1989).* More than 200 book chapters, journal articles, and books describe how CBM can be used by teachers and other educators to make important decisions about students' instructional needs.

Four critical features distinguish CBM:

1. Students are tested on an ongoing basis from multiple reading tests,
2. Tests are standardized, short tasks.
3. Tests measure an important key skill, (oral reading).
4. Tests use reading passages of about equal difficulty.

Originally, the source of students' reading passages for CBM was their own general education curriculum. If the basal series was Houghton Mifflin, teachers would prepare their own reading passages using systematic samples from Houghton Mifflin that were retyped, words were counted, and copies made and distributed.

Over time, it became apparent that, more often than not, teachers and schools, and of course, school districts and states, differed radically in their choice of basal reading series. Equally as frequent, many teachers and schools did not use a basal text at all, relying on trade books or children's literature as the source of their instructional materials. Additionally, it became apparent that even when a common basal series was used, the stories within each book (Book 3-2) often varied in difficulty more than the stories between books (Book 3-2 and Book 4). Finally, teachers were spending too much time creating materials that varied in quality.

In sum, when schools created passages from their own reading series, it was possible to achieve high integrity on CBM Feature 1 (ongoing testing) and CBM Feature 3 (measuring oral reading). However, the variability in text difficulty within- and between-books and series raised concerns about the integrity of CBM Feature 2 (standard tasks) and CBM Feature 4 (passages of approximately equal difficulty).

The technical challenge became, then, not what to measure or how often, but the development of reading passages that would be (a) standard, (b) both curriculum-independent and sensitive to all reading curriculum, (c) and of about equal difficulty within each grade-level. The question was whether using standard, graded passages that were "curriculum independent," could be used to draw the same conclusions about students' reading status compared to other readers at one point in time and their rates of progress over time.

An ongoing program of research has established that the key feature(s) of assessment procedures like CBM is not specifically the use of students' own reading curriculum passages (Fuchs & Deno, 1994; Tilly, 1999). Although passages taken directly from a student's own curriculum may have high "instructional validity," the challenges of feasibility (e.g., developing the passages) as well as the more important issues of consistency and comparability of decision making based on passages of wide variability demanded a different approach. That approach was to assess students on a repeated basis using standard, more "generic" or "curriculum-independent" graded passages. The challenge, however, was to develop a sufficient number of graded passages within each grade that could be used with confidence to make decisions about students' general reading skills and progress.

Creating Standard Graded Reading Passages of Equal Difficulty

To accomplish this task, standard Reading Assessment Passages (RAPs) were developed for Grades 1 through 8 to be used with the ***AIMSweb Systems***. These Reading Assessment Passages (RAPs) were written based on the following general principles

1. The passages had to be authored by persons familiar with the teaching of reading and how students learn to read. These persons also had to be familiar with the kinds of literature that students at the various grades encountered.

2. The passages had to be written to fit initially within the parameters established by a common formula used to calculate readability. Follow-up testing had to be conducted to look at the generalizability of the readability, or difficulty ratings to other commonly used readability formulas.

Recruitment of Authors and Creating Passages for Initial Fit

A written notice describing the search for authors to develop graded reading passages was sent to all educators in a medium-size suburban/rural education district in the Midwest of approximately 6,100 students. Thirty teachers and 10 paraprofessionals who were interested met and were given (a) a presentation about the purpose of the project, (b) a description of the measurement system to be used to pilot the RAPs, (c) sample passages, and (d) a written copy of guidelines for creating passages at each grade level. The guidelines given to authors included:

1. The number of words to be written for each passage by grade, and,

2. Criteria for the number of syllables and sentences per 100 words per grade as determined by the Fry (1968) readability formula.

Authors were be asked to write passages of 250 words for first and second grade, 300 words for third grade, and 350 words for Grades 4 through 8. Writers also practiced the use of the Fry readability formula, by determining the grade level of at least five, 100-word selections. In addition, authors were given instruction on how to write proper names witihin the passages, and were asked to write a story with a beginning and ending.

Following the initial training, 9 teachers and 7 paraprofessionals produced passages for field-testing. Once an author completed a story, they sent the story to the first author for review. These passages were either (a) edited and given back to the writer to finish final editing, or (b) rejected due to inappropriate subject matter. Authors continued to produce passages until a pool of 33 were written for 1st Grade 1 and 50 were written for Grades 2-8.

Selecting Standard Graded Reading Assessment Passages of About Equal Difficulty

The initial pool of passages was reduced to a final set of 23 passages for Grade 1 and 33 passages for Grades 2-8 through examining actual student oral reading scores.

1. A sample of students at each grade read all passages developed for their grade level, and

2. Passages were dropped that:
   a. Had low alternate-form reliability, and
   b. Means, standard deviations (SD), and standard errors of measurement (SEM) showed the most variabliity in means, standard deviations, and standard errors of measurement (SEM) by grade.

Procedures for Collecting Student Reading Data

Twenty-four students per grade, first through eighth, were chosen from 4 elementary and 3 middle schools in a suburban/rural mid-western education district. At each of the 4 schools, 6 students from each grade were selected based on this sampling plan:

2 students from the 75th percentile,
2 from the 50th percentile, and
2 from the 25th percentile.

The sixth-grade group included only 18 students because of scheduling difficulties. All RAP field-testing was completed in February and March of 2001.

Trained examiners conducted the testing during a 4-week period from mid-February to mid-March, 2001. Each student read each passage from the pool of passages for the grade for 1-minute and the number of words read correct (WRC) and errors were calculated. Testing was divided up into 5 sessions. Each student grades 2-8 read approximately 10 passages per session. Students in first grade read 6-7 passages per session.

Procedures for Examining Student Performance Data

Once student data were collected, the following procedures were used to select a final pool of passages to be used with the ***AIMSweb Systems***.

Alternate-Form Reliability

Equal difficulty of passages had to be established empirically through calculation of alternate-form reliabilities. To calculate alternate-form reliability, all possible pairs of students' passage scores were correlated within each grade. Generally, those passages not highly correlated (greater than .70) with other passages within the same grade were discarded.

Mean reliability correlations then were calculated for each passage. An overall reliability coefficient for the passages at each grade was calculated by taking the average of the mean correlations for each passage. This overall reliability coefficient was subsequently used to calculate the standard error of measurement (SEM) for passages at each grade.

Comparing Means, Standard Deviations, and Standard Errors of Measurement
At each grade, students' scores (number of Words Read Correct (WRC)) on each passage were averaged to create a passage mean WRC and standard deviation. Passage means then were averaged to calculate an overall mean and standard deviation for each grade. The standard deviation and reliability coefficient for each grade were then used to calculate the SEM for each grade. Passages whose mean WRC were more than +1.0 standard error of measurement (SEM) outside the grade-level mean WRC then were discarded. For example, if the grade-level mean was 125 WRC and the SEM=11, passages with scores of 137 WRC or more and scores off 113 WRC or less, were discarded.

Readability Comparisons
Passages also were examined using the Lexile-graded standards (Stenner & Burdick, 1997), a method of estimating reading passage difficulty. Passages that did not receive Lexile scores within a range acceptable for their grade for were eliminated from the pool.

Selecting a Final Pool of Passages

Upon examination of student performance data using the 3 methods described above, a final pool of 23 passages for Grade 1 and 33 passages for Grades 2-7 remained. Passages written for Grade 8 produced means and standard deviations that were lower overall than means and standard deviations for Grade 7. Therefore, it was decided that the 8th grade passages were too difficult and variable and would need to be re-written. Additionally, more than half of the passages produced alternate-form reliability scores below .70.

Passages for Grade 8 were then edited by the second author and field-tested in April-May 2002. Fourteen schools (n=183 students) from 5 states were mailed 50 passages along with directions for adminstration and scoring. Each school trained examiners to administer and score the passages. Approximately 10 to 13 randomly selected students from each school read 10 randomly selected passages from the pool of 50. Scores were analyzed across schools using the same procedures utilized for selected passages for Grades 1-7. The average number of students to read a passage was 36, with a range of 19 to 55. From the pool at each grade, 3 passages were selected for use in ***AIMSweb Benchmark.*** These passages produced:

1. Similar means in WRC and standard deviations,
2. Had high alternate form reliability (> .70), and
3. Had appropriate Lexile scores.

The remainder of the passages were selected as the passages for use with ***AIMSweb Progress Monitor.*** Because students' reading is tested more frequently in this approach, error variability attributable to differences in passage difficulty is reduced.

Finally, comparisons of the 3 *Benchmark RAPs* and *Progress Monitoring RAPs* to other commonly accepted readability formulas were made. Correlations were calculated between passages, Lexile scores, and the following readability formulas: Dale-Chall, Flesh, Powers-Sumner-Kearl, SMOG, and the Spache as they are the most appropriate formulas to use for elementary and secondary grade levels. Readability correlations were calculated by using *Readability Calculations* (1999), a software program developed by Micro Power & Light Co. to compare multiple formulae.

Technical Features of AIMSweb RAPs

The technical features of the *AIMSweb Standard Reading Assessment Passages* are organized at three levels:
1. Technical features of all RAPs within a grade,
2. Technical features of the 3 RAPs used with ***AIMSweb Benchmark*** by grade, and,
3. Technical features of the 20-30 RAPs used with ***AIMSweb Progress Monitor*** passages by grade.

Overall Technical Features of RAPs by Grade

Table 1 contains the overall means, standard deviations, standard errors of measurement, and reliability coefficients for passages by grade. Developmental trends, as expected, confirmed an increase in mean WRC across grades. The mean is lower and the standard deviation larger for Grade 8 passages. However, passages at the extremes were omitted as they produced low alternate-form reliability scores and considered less reliable. Therefore, the final set of Grade 8 passages is considered more reliable even though the overall mean is lower than Grade 7 passages. Standard deviations were consistent with other published studies. SEMs were similar for Grades 2-8 ranging from 6.3 to 13.3 The SEM was smaller for Grade 1 as there were 10 less passages selected. There was high-alternate form reliability across grades ranging from .81 to .90.

Table 1. Overall Technical Features of *AIMSweb Standard Reading Assessment Passages* By Grade

Grade	1	2	3	4	5	6	7	8
n	23	33	33	33	33	33	33	33
Mean Words Read Correct (WRC)[a]	35.7	89.8	107.6	121.5	132.1	141.8	154.2	147.3
Standard Deviation[b]	19.2	23.6	28.1	25.3	29.1	25.1	25.0	42.0
Standard Error of Measurement (SEM)	6.3	9.9	10.5	9.7	10.5	10.0	11.2	13.3
Reliability[c]	.89	.83	.86	.85	.88	.84	.80	.90

Notes:
n= number of passages
a= Average of RAP grade means
b= Average of RAP grade standard deviations
c= Average of RAP grade correlations (alternate-form reliability)

Overall Technical features of the 3 RAPs used with AIMSweb Benchmark testing

Technical features for the 3 RAPs used with *AIMSweb Benchmark* testing are displayed in Table 2. For each passage, the mean WRC, standard deviation, alternate-form reliability, and Lexile score are displayed. The mean WRC, standard deviation, and alternate-for reliability of the 3 Benchmark RAPs are displayed by grade.

Means were highly similar for each grade with a maximum difference of about 5.8 WRC per minute within the grade. Alternate-form reliability correlations were all greater than .80 with the exception of Passage 1 at Grade 7. Thirteen of 21 passages exceeded .85. As the grade level of the passages

increased, so did the Lexile scores for these passages. This correspondence confirmed the expected increases in difficulty from Grade 1 passages to Grade 7 passages.

Table 2. Technical Features of *AIMSweb Standard Benchmark Reading Assessment Passages*

Grade	Passage	Mean WRC	Standard Deviation	Alternate Form Reliability[b]	Lexile
1	1	37.0	18.6	0.91	240
	2	35.8	18.7	0.91	210
	3	35.9	18.8	0.89	250
	Mean	**36.2**	**18.7**	**0.90**	
2	1	89.6	20.8	0.81	420
	2	89.1	19.7	0.80	440
	3	89.2	23.6	0.85	470
	Mean	**89.3**	**21.4**	**0.82**	
3	1	107.1	29.4	0.85	630
	2	107.2	31.7	0.83	460
	3	109.1	28.2	0.87	570
	Mean	**107.8**	**29.8**	**0.85**	
4	1	121.5	20.1	0.87	770
	2	121.8	27.2	0.82	650
	3	122.8	24.5	0.86	670
	Mean	**122.0**	**23.9**	**0.85**	
5	1	133.0	29.1	0.85	810
	2	131.8	29.7	0.89	780
	3	136.3	29.3	0.89	770
	Mean	**133.7**	**29.4**	**0.88**	
6	1	141.1	27.5	0.82	1050
	2	143.8	23.6	0.85	840
	3	138.0	26.5	0.87	930
	Mean	**141.0**	**25.8**	**0.85**	
7	1	154.6	27.6	0.73	1060
	2	153.7	21.4	0.83	970
	3	154.0	22.0	0.82	1000
	Mean	**154.1**	**23.7**	**0.79**	
8	1	149.5	42.0	.90	1040
	2	146.7	47.3	.91	1070
	3	145.5	35.7	.94	1010
	Mean	**147.2**	**41.7**	**.92**	

Notes:
a= Standard error of measurement (SEM) for all RAP
b= Mean correlation for each alternate-form RAP

Comparisons of passages and Lexile scores for the 3 Benchmark RAPSs with selected readability formulas are displayed in Table 3. Because readability data are ordinal, each set of graded passages was assigned a number for that grade to calculate correlations across all grades (n=254 passages). Thus, the 3 Benchmark RAPs at Grade 2 were assigned a value of 2 while the passages at Grade 5 were assigned a value of 5. Correlations were high ranging between .78 and .99 with a median correlation of .95.

Table 3. Readability Correlations for *AIMSweb Standard Benchmark Reading Assessment Passages*

	Passage Level	Lexile	Fry	Flesch	Powers	Spache
Passage Level						
Lexile	0.97					
Fry	0.94	0.90				
Flesch	0.96	0.92	0.99			
Powers	0.92	0.88	0.98	0.98		
Spache	0.97	0.93	0.95	0.97	0.94	
SMOG	0.83	0.78	0.92	0.91	0.95	0.86
Median	**0.95**					

Comparisons of passages and Lexile scores for the Progress Monitoring RAPSs with selected readability formulas are displayed in Table 4. Again, because readability data are ordinal, each set of graded passages was assigned a number for that grade to calculate correlations. Thus, the 30 Progress Monitoring RAPs at Grade 2 were assigned a value of 2 while the passages at Grade 5 were assigned a value of 5. Correlations were moderate to high ranging between .78 and .98. with a median correlation of .90.

Table 4. Readability Correlations for *AIMSweb Standard Progress Monitor Reading Assessment Passages*

	Passage Level	Lexile	Fry	Flesch	Powers	Spache
Passage Level						
Lexile	0.90					
Fry	0.91	0.88				
Flesch	0.88	0.87	0.98			
Powers	0.89	0.85	0.98	0.98		
Spache	0.81	0.83	0.91	0.95	0.90	
SMOG	0.84	0.78	0.92	0.92	0.94	0.84
Median	**0.90**					

Overall Technical Features of AIMSweb RAPs

The technical features of *AIMSweb* RAPs by grade are displayed in Tables 5-11, including mean WRC, standard deviation, alternate-form reliability, and Lexile score. For each grade, mean WRC are sorted from the highest number of WRC across students to the lowest number of WRC. Passages numbered 1, 2, and 3 and in bold are the 3 Benchmark RAPs for each grade. If a cell is empty in the Lexile score column, no score was calculated.

Across grades, mean WRC are within +1 SEM of the overall mean. Standard deviations are similar across grades. Alternate-form reliability correlations were high across grades with 237 of 254 passages above .80. Lexile scores increased as the difficulty of passages increased across grades .

Table 5. Grade 1 Technical Features (n=23)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
4	42.1	22.3	0.90	360
19	40.6	20.3	0.89	190
12	40.5	23.5	0.91	360
16	39.6	18.1	0.86	300
22	38.6	22.2	0.84	250
23	38.5	21.7	0.89	400
10	37.9	19.7	0.90	190
14	37.5	19.2	0.89	200
1	**37.0**	**18.6**	**0.91**	**240**
18	36.8	18.1	0.88	260
17	36.3	17.7	0.91	290
3	**35.9**	**18.8**	**0.89**	**250**
2	**35.8**	**18.7**	**0.91**	**210**
6	35.0	14.5	0.82	390
11	35.0	22.7	0.90	220
20	34.1	15.2	0.90	300
15	34.0	20.8	0.90	280
8	32.7	17.2	0.90	310
9	32.2	20.3	0.92	260
21	30.9	15.9	0.88	410
13	30.5	19.4	0.93	410
5	30.1	16.6	0.85	370
7	29.0	19.5	0.91	200
Mean	**35.7**	**19.2**	**0.89**	

Table 6. Grade 2 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
31	99.9	24.2	0.87	220
7	99.6	25.5	0.87	370
9	99.0	22.8	0.82	340
18	98.9	20.9	0.83	470
20	98.0	26.8	0.83	400
4	97.9	24.3	0.81	340
30	97.1	23.9	0.82	300
27	95.4	21.8	0.81	380
11	94.6	21.9	0.84	440
8	94.5	21.9	0.87	510
28	93.6	24.3	0.84	310
15	93.0	23.9	0.83	490
10	91.7	16.9	0.76	520
24	91.5	23.2	0.80	450
23	90.5	29.3	0.89	460
29	90.3	22.1	0.81	260
1	**89.6**	**20.8**	**0.81**	**420**
14	89.6	30.3	0.87	300
3	**89.2**	**23.6**	**0.85**	**470**
2	**89.1**	**19.7**	**0.80**	**440**
33	87.5	24.3	0.82	340
12	85.8	21.8	0.84	540
13	85.2	23.5	0.81	450
6	84.3	24.5	0.83	500
16	83.7	24.9	0.86	320
19	83.3	24.8	0.83	420
26	82.6	24.8	0.83	290
22	82.5	24.7	0.80	490
32	82.0	22.2	0.86	500
21	81.7	29.8	0.87	400
5	81.6	19.9	0.79	530
17	80.3	23.9	0.86	440
25	80.0	22.6	0.87	370
Mean	**89.8**	**23.6**	**0.83**	

Table 7. Grade 3 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
17	113.0	27.2	0.87	540
19	112.8	28.2	0.84	610
28	112.4	24.0	0.85	470
26	112.0	29.7	0.86	620
6	112.0	29.8	0.87	500
7	111.8	27.6	0.84	520
27	110.5	28.2	0.85	710
9	110.4	26.3	0.86	590
31	110.2	23.2	0.87	640
32	109.7	27.4	0.90	600
20	109.6	23.4	0.85	640
3	**109.1**	**28.2**	**0.87**	**570**
14	109.0	31.0	0.84	430
12	108.7	31.4	0.85	660
2	**107.2**	**31.7**	**0.83**	**460**
1	**107.1**	**29.4**	**0.85**	**630**
23	106.8	24.1	0.83	690
13	106.8	28.2	0.87	590
33	106.1	30.8	0.87	440
8	106.0	26.7	0.84	570
15	105.7	32.0	0.88	520
11	105.5	29.9	0.87	570
25	105.5	28.7	0.86	520
16	105.4	25.1	0.83	600
30	105.4	29.7	0.85	630
10	105.1	30.9	0.89	580
24	105.0	33.6	0.85	600
21	104.6	30.0	0.83	580
18	104.5	27.4	0.87	500
5	103.8	27.7	0.86	520
22	103.8	26.0	0.83	620
29	103.1	23.2	0.89	660
4	102.9	27.6	0.88	640
Mean	**107.6**	**28.1**	**0.86**	

Table 8. Grade 4 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
29	129.5	23.1	0.85	400
8	129.4	26.6	0.87	750
6	129.0	32.2	0.89	610
32	128.6	27.4	0.87	690
11	128.4	29.4	0.85	560
30	127.8	27.3	0.84	760
26	127.0	20.8	0.83	560
14	126.8	18.8	0.75	680
22	125.4	25.9	0.87	600
20	125.1	29.8	0.88	630
33	125.1	29.9	0.89	600
19	124.9	24.4	0.84	700
17	123.2	28.4	0.88	610
10	123.0	21.9	0.86	490
3	**122.8**	**24.5**	**0.86**	**670**
21	122.8	29.5	0.86	480
24	122.5	19.9	0.86	620
2	**121.8**	**27.2**	**0.82**	**650**
1	**121.5**	**20.1**	**0.87**	**770**
23	121.0	25.4	0.85	670
4	119.6	28.7	0.90	500
28	119.2	24.3	0.82	610
12	118.3	22.9	0.85	530
13	117.7	18.9	0.82	550
5	117.4	25.2	0.85	580
7	117.2	22.6	0.83	720
31	116.3	26.5	0.85	720
9	116.2	22.1	0.85	580
25	115.3	20.4	0.87	690
15	115.0	22.7	0.84	580
27	112.1	28.1	0.87	590
18	111.6	29.6	0.86	620
16	109.1	28.8	0.84	750
Mean	**121.5**	**25.3**	**0.85**	

Table 9. Grade 5 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
9	141.7	29.0	0.87	790
22	141.5	30.1	0.89	780
11	141.5	32.7	0.90	710
13	141.3	35.4	0.87	560
27	140.2	28.7	0.88	690
14	140.0	28.7	0.89	1010
10	138.3	31.6	0.87	650
33	137.4	29.6	0.86	570
28	137.1	30.1	0.91	720
3	**136.3**	**29.3**	**0.89**	**770**
26	135.5	27.7	0.87	820
23	134.8	28.3	0.87	440
31	133.8	32.2	0.90	610
30	133.7	27.0	0.87	650
6	133.6	30.7	0.87	580
18	133.6	24.0	0.86	650
1	**133.0**	**29.1**	**0.85**	**810**
2	**131.8**	**29.7**	**0.89**	**780**
24	131.5	32.2	0.86	630
15	130.3	29.2	0.86	560
21	129.5	34.8	0.88	830
5	128.2	20.8	0.87	780
8	128.2	29.9	0.89	930
12	128.0	29.0	0.87	870
7	126.7	29.8	0.83	740
19	125.8	27.2	0.91	760
17	125.6	27.4	0.86	790
25	125.0	27.1	0.85	610
4	123.8	28.1	0.91	660
32	123.7	32.2	0.91	
16	123.0	24.1	0.91	750
29	122.8	27.2	0.88	660
20	121.6	27.6	0.88	880
Mean	**132.1**	**29.1**	**0.88**	

Table 10. Grade 6 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
10	153.6	30.3	0.78	700
29	152.9	25.6	0.85	560
9	151.4	22.0	0.81	610
4	150.4	24.3	0.79	570
19	148.9	28.4	0.84	940
31	147.3	24.7	0.86	710
25	146.8	23.6	0.86	520
18	146.6	23.7	0.85	810
26	146.0	28.9	0.88	810
17	145.7	28.7	0.87	880
32	145.4	22.9	0.81	590
23	144.0	27.6	0.85	730
2	**143.8**	**23.6**	**0.85**	**840**
7	143.4	28.3	0.86	1000
12	142.7	26.5	0.85	920
6	142.2	22.6	0.82	740
5	141.8	22.2	0.84	630
24	141.7	20.2	0.85	840
1	**141.1**	**27.5**	**0.82**	**1050**
8	141.1	23.1	0.84	740
14	140.6	26.5	0.88	790
33	140.6	27.8	0.82	690
15	140.1	17.1	0.71	1030
30	139.8	28.2	0.85	810
11	139.0	29.4	0.87	780
3	**138.0**	**26.5**	**0.87**	**930**
21	136.7	23.1	0.82	720
20	134.4	25.0	0.88	850
27	134.0	25.7	0.85	740
22	132.8	19.5	0.78	680
28	132.1	24.4	0.87	830
16	131.5	24.9	0.85	770
13	124.6	24.9	0.85	880
Mean	**141.8**	**25.1**	**0.84**	

Table 11. Grade 7 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
30	165.0	24.5	0.84	650
31	163.4	23.9	0.80	880
8	163.3	26.8	0.85	990
20	163.3	20.1	0.78	910
16	162.7	23.9	0.81	900
28	162.6	30.7	0.84	830
19	162.6	26.8	0.82	810
13	162.0	29.5	0.80	780
32	161.3	20.9	0.80	870
25	159.9	24.5	0.81	860
7	159.7	26.7	0.82	970
24	159.3	29.4	0.69	950
29	158.0	29.4	0.75	720
15	157.4	25.0	0.81	1010
6	156.9	21.0	0.84	950
26	156.7	22.8	0.82	880
10	154.9	26.4	0.80	910
18	154.7	24.1	0.81	760
1	**154.6**	**27.6**	**0.73**	**1060**
3	**154.0**	**22.0**	**0.82**	**1000**
2	**153.7**	**21.4**	**0.83**	**970**
27	153.1	22.1	0.73	900
33	151.8	20.7	0.77	910
5	151.1	29.2	0.86	730
22	149.2	23.5	0.80	870
21	149.0	22.4	0.76	870
23	148.6	24.0	0.82	840
4	145.9	28.1	0.82	670
17	143.3	20.9	0.81	860
11	142.5	23.9	0.85	990
14	140.1	26.4	0.86	810
12	136.5	25.6	0.79	830
9	134.2	25.8	0.75	780
Mean	**154.2**	**25.0**	**0.80**	

Table 12. Grade 8 Technical Features (n=33)

Passage Number	Mean WRC	Standard Deviation	Alternate-Form Reliability	Lexile Score
27	161.88	44.19	0.93	880
25	160.52	49.13	0.93	820
6	**158.28**	**41.99**	**0.86**	**910**
8	157.73	43.82	0.89	1090
10	156.41	36.78	0.90	800
16	156.18	40.65	0.89	990
24	156.02	48.22	0.93	930
29	155.56	47.34	0.93	900
14	152.57	44.31	0.91	860
21	150.45	42.80	0.84	1100
1	**149.49**	**41.98**	**0.90**	**1040**
18	147.83	41.13	0.81	940
5	147.53	39.08	0.93	850
2	**146.70**	**47.29**	**0.91**	**1070**
3	**145.52**	**35.73**	**0.94**	**1010**
31	145.40	33.50	0.83	1170
23	145.30	35.63	0.89	1130
15	145.26	47.21	0.93	920
12	145.11	48.55	0.92	980
22	144.58	36.37	0.88	1100
9	143.75	40.49	0.91	1090
11	143.62	37.66	0.82	980
13	143.57	48.09	0.94	1000
26	143.40	42.38	0.95	1010
32	143.13	49.47	0.77	1140
7	141.26	39.44	0.87	1030
28	140.98	39.90	0.83	990
33	140.31	41.52	0.90	970
30	140.20	49.47	0.89	1020
19	139.82	40.56	0.93	1090
17	138.43	40.63	0.92	1170
20	137.72	39.36	0.95	1080
4	137.50	34.06	0.92	960
Mean	**147.33**	**42.08**	**0.90**	

Summary

In sum, the *AIMSweb Standard Reading Assessment Passages* for Grades 1 through 8 are technically adequate and of sufficient equal difficulty to be used with confidence to make decisions about students' general reading skills and progress. The RAPs can be used with confidence to assess student reading performance and overcome the inherent difficulties associated with using a student's "own" curriculum.

References

Deno, S.L. (1985). *Curriculum-based measurement: The emerging alternative. Exceptional Children*, 52, (3), pp 219-232.

Fuchs, L.S. and Deno, S. L. (1994). *Must instructionally useful performance assessment be based in the curriculum? Exceptional Children*, 61, (1), pp 15-24.

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, 11, pp 513-16.

Readability Calculations (1999). Micro Power & Light Co. Dallas: TX.

Shinn, M.R.(Ed.) (1989). *Curriculum-Based Measurement: Assessing Special Children.* Guilford School Practitioner Series: New York.

Stenner, A. J., Smith, M., and Burdick, D. S. (1983). *Toward a theory of construct definition. Journal of Educational Measurement*, 20, 305-315.

Tilly, D.W. (1999). *The effect of passage difficulty on reading curriculum-based measurement data: A generalizable theory approach.* Presentation at Pacific Coast Research Council. La Jolla, California.