



I410L | Lexile: Matching readers to text

# Linking the AIMS with the Lexile Framework

## **MetaMetrics, Inc.**

1000 Park Forty Plaza Drive, Suite 120

Durham, NC 27713

[www.Lexile.com](http://www.Lexile.com)

August 31, 2007

# Linking the *AIMS* with the Lexile Framework

A Study to Link  
the Arizona Instrument to Measure Standards Reading Scale  
with The Lexile Framework<sup>®</sup> for Reading

Prepared by MetaMetrics, Inc. for the Arizona Department of Education, Standards and Assessment Division, 1535 West Jefferson Street, Phoenix, Arizona 85007

MetaMetrics, Inc.  
1000 Park Forty Plaza Drive, Suite 120  
Durham, North Carolina 27713  
[www.lexile.com](http://www.lexile.com)

Final Report, August 31, 2007

## Introduction

Often it is desirable to convey more information about test performance than can be incorporated into a single primary score scale. Two examples of this arise in large-scale assessment. In one situation, one test can provide a unique type of information (such as national comparisons available from NAEP) but is not administered very often. At the same time another test is administered more often, but is not able to provide the breadth of information (such as a state assessment). An auxiliary score scale for a test can be established to provide this additional information through assessment scale linkages. Once linkages are established between the two assessments, then the results of the more-frequently-administered assessment can be translated in terms of the scale for the other assessment.

In another situation, the linkage between two score scales can be used to provide a context for understanding the results of one of the assessments. Sometimes it is hard to explain what a student can read based on the results of a reading comprehension test. Parents typically ask the questions “If my child scores 456 on Arizona’s Instrument to Measure Standards Dual Purpose Assessment (AIMS DPA), what does this mean?” or “Based on my child’s test results, what can he or she read and how well?” Once a linkage is established with an assessment that is related to specific titles, then the results of the assessment can be explained and interpreted in the context of the specific titles that a student can read.

Auxiliary score scales can be used to “convey additional normative information, test-content information, and information that is jointly normative and content based. For many test uses, an auxiliary scale conveys information that is more crucial than the information conveyed by the primary score scale. In such instances, the auxiliary score is the one that is focused on, and the primary scale can be viewed more as a vehicle for maintaining interpretability over time” (Petersen, Kolen, and Hoover, 1989, p. 222). One such auxiliary scale is The Lexile Framework<sup>®</sup> for Reading, which was developed to appropriately match readers with text at a level that provides challenge but not frustration.

Linking assessment results to the Lexile Framework provides a mechanism for matching each student with text on a common scale. It serves as an anchor to which texts and assessments can be connected allowing parents, teachers, and administrators to speak the same language. By using the Lexile Framework, the same metric is applied to the books the children read, the tests they take, and the results that are reported. Parents often ask questions like the following:

- How can I help my child become a better reader?
- How do I challenge my child to read?

Questions like these can be challenging for parents and educators. By linking the Arizona state assessments with the Lexile Framework, educators and parents will be able to answer these questions and will be better able to use the results from the test to improve instruction and to develop each student’s level of reading comprehension.

This research study was designed to provide a mechanism to provide reading levels that can be matched to text based on Arizona’s Instrument to Measure Standards Dual Purpose Assessment (AIMS DPA) and the Arizona’s Instrument to Measure Standards High School (AIMS HS) results. The study was funded by the Arizona Department of Education (Purchase

Order No. E7PD2986) and conducted by the staff of MetaMetrics, Inc. and in collaboration with the Arizona Department of Education. The primary purposes of this study were:

- to provide the Arizona Department of Education with tools that could be used to answer questions related to standards, student-level accountability, test score interpretation, and test validation;
- to provide conversion tables for predicting Lexile measures from the Arizona's Instrument to Measure Standards Dual Purpose Assessment (Grades 3 through 8) and Arizona's Instrument to Measure Standards High School (Second Year in High School) test results; and
- to produce a report that describes the linking analysis procedures.

# The Lexile Framework for Reading

All symbol systems share two features: a semantic component and a syntactic component. In language, the semantic units are words. Words are organized according to rules of syntax into thought units and sentences (Carver, 1974). In all cases, the semantic units vary in familiarity and the syntactic structures vary in complexity. The comprehensibility or difficulty of a message is dominated by the familiarity of the semantic units and by the complexity of the syntactic structures used in constructing the message.

## The Semantic Component

As far as the semantic component is concerned, it is clear that most operationalizations are proxies for the probability that an individual will encounter a word in a familiar context and thus be able to infer its meaning (Bormuth, 1966). This is the basis of exposure theory, which explains the way receptive or hearing vocabulary develops (Miller and Gildea, 1987; Stenner, Smith, and Burdick, 1983). Klare (1963) hypothesized that the semantic component varied along a familiarity-to-rarity continuum. This concept was further developed by Carroll, Davies, and Richman (1971), whose word-frequency study examined the reoccurrence of words in a five-million-word corpus of running text. Knowing the frequency of words as they are used in written and oral communication provided the best means of inferring the likelihood that a word would be encountered by a reader and thus become a part of that individual's receptive vocabulary.

Variables such as the average number of letters or syllables per word have been observed to be proxies for word frequency. There is a strong negative correlation between the length of words and the frequency of word usage. Polysyllabic words are used less frequently than monosyllabic words, making word length a good proxy for the likelihood that an individual will be exposed to a word.

In a study examining receptive vocabulary, Stenner, Smith, and Burdick (1983) analyzed more than 50 semantic variables in order to identify those elements that contributed to the difficulty of the 350 vocabulary items on Forms L and M of the *Peabody Picture Vocabulary Test—Revised* (Dunn and Dunn, 1981). Variables included part of speech, number of letters, number of syllables, the modal grade at which the word appeared in school materials, content classification of the word, the frequency of the word from two different word counts, and various algebraic transformations of these measures.

The first word frequency measure used was the raw count of how often a given word appeared in a corpus of 5,088,721 words sampled from a broad range of school materials (Carroll, Davies, and Richman, 1971). For example, the word "accident" appears 176 times in the 5,088,721-word corpus. The second word frequency measure used was the frequency of the "word family." A word family included: (1) the stimulus word; (2) all plurals (adding "-s" or "-es" or changing "-y" to "-ies"); (3) adverbial forms; (4) comparatives and superlatives; (5) verb forms ("-s," "-d," "-ed," and "-ing"); (6) past participles; and (7) adjective forms. For example, the word family for "accident" would include "accidental," "accidentally," "accidentals," and "accidents," and they would all have the same word frequency of 334. The frequency of a word family was based on the sum of the individual word frequencies from each of the types listed.

Correlations were computed between algebraic transformations of these means (mean frequency of the words in the test item and mean frequency of the word families in the test item) and the rank order of the test items. Since the items were ordered according to increasing difficulty, the rank order was used as the observed item difficulty. The log of the mean word frequency provided the strongest correlation with item rank order ( $r = -0.779$ ) for the items on the combined form.

The Lexile Framework currently employs a 300-million-word corpus when examining the semantic component of text. This corpus was assembled from the more than 15,000 texts that were measured by MetaMetrics for publishers from 1998 through January 2000. When text is analyzed by MetaMetrics, all electronic files are initially edited according to established guidelines used with the Lexile Analyzer software. These guidelines include the removal of all incomplete sentences, chapter titles, and paragraph headings; running of a spell check; and re-punctuating where necessary to correspond to how the book would be read by a child (for example, at the end of a page). The text is then submitted to the Lexile Analyzer that examines the lengths of the sentences and the frequencies of the words and reports a Lexile measure for the book. When enough additional texts have been analyzed to make an adjustment to the corpus necessary and desirable, a linking study will be conducted to adjust the calibration equation such that the Lexile measure of a text based on the current corpus will be equivalent to the Lexile measure based on the new corpus.

## **The Syntactic Component**

Klare (1963) provides a possible interpretation for how sentence length works in predicting passage difficulty. He speculated that the syntactic component varied with the load placed on short-term memory. Crain and Shankweiler (1988), Shankweiler and Crain (1986), and Liberman, Mann, Shankweiler, and Westelman (1982) have also supported this explanation. The work of these individuals has provided evidence that sentence length is a good proxy for the demand that structural complexity places upon verbal short-term memory.

While sentence length has been shown to be a powerful proxy for the syntactic complexity of a passage, an important caveat is that sentence length is not the underlying causal influence (Chall, 1988). Researchers sometimes incorrectly assume that manipulation of sentence length will have a predictable effect on passage difficulty. Davidson and Kantor (1982), for example, illustrated rather clearly that sentence length can be reduced and difficulty increased and vice versa.

Based on previous research, it was decided to use sentence length as a proxy for the syntactic component of reading difficulty in the Lexile Framework.

## **Calibration of Text Difficulty**

The research study on semantic units (Stenner, Smith, and Burdick, 1983) was extended to examine the relationship of word frequency and sentence length to reading comprehension. In 1987(a), Stenner, Smith, Horiban, and Smith performed exploratory regression analyses to test the explanatory power of these variables. This analysis involved calculating the mean word frequency and the log of the mean sentence length for each of the 66 reading comprehension passages on the *Peabody Individual Achievement Test* (Dunn and Markwardt, 1970). The observed difficulty of each passage was the mean difficulty of the items associated with the

passage (provided by the publisher) converted to the logit scale. A regression analysis based on the word-frequency and sentence-length measures produced a regression equation that explained most of the variance found in the set of reading comprehension tasks. The resulting correlation between the observed logit difficulties and the theoretical calibrations was 0.97 after correction for range restriction and measurement error. The regression equation was further refined based on its use in predicting the observed difficulty of the reading comprehension passages on 8 other standardized tests. The resulting correlation between the observed logit difficulties and the theoretical calibrations across the 9 tests was 0.93 after correction for range restriction and measurement error.

Once a regression equation is established linking the syntactic and semantic features of text to the difficulty of text, then the equation can be used to calibrate test items and text.

## The Lexile Scale

In developing the Lexile Scale, the Rasch model (Wright and Stone, 1979) was used to estimate the difficulties of the items and the abilities of the persons on the logit scale.

The calibrations of the items from the Rasch model are objective in the sense that the relative difficulties of the items will remain the same across different samples of persons (specific objectivity). When two items are administered to the same group it can be determined which item is harder and which one is easier. This ordering should hold when the same two items are administered to a second group. If two different items are administered to the second group, there is no way to know which set of items is harder and which set is easier. The problem is that the location of the scale is not known. General objectivity requires that scores obtained from different test administrations be tied to a common zero—absolute location must be sample independent (Stenner, 1990). To achieve general objectivity, the theoretical logit difficulties must be transformed to a scale where the ambiguity regarding the location of zero is resolved.

The first step in developing a scale with a fixed zero was to identify two anchor points for the scale. The following criteria were used to select the two anchor points: they should be intuitive, easily reproduced, and widely recognized. For example, with most thermometers the anchor points are the freezing and boiling points of water. For the Lexile Scale, the anchor points are text from seven basal primers for the low end and text from *The Electronic Encyclopedia* (Grolier, Inc., 1986) for the high end. These points correspond to the middle of first grade text and the midpoint of workplace text.

The next step was to determine the unit size for the scale. For the Celsius thermometer, the unit size (a degree) is 1/100<sup>th</sup> of the difference between freezing (0 degrees) and boiling (100 degrees) water. For the Lexile Scale the unit size (a Lexile) was defined as 1/1000<sup>th</sup> of the difference between the mean difficulty of the primer material and the mean difficulty of the encyclopedia samples. Therefore, a Lexile by definition equals 1/1000<sup>th</sup> of the difference between the difficulty of the primers and the difficulty of the encyclopedia.

The third step was to assign a value to the lower anchor point. The low-end anchor on the Lexile Scale was assigned a value of 200.

Finally, a linear equation of the form

$$[(\text{Logit} + \text{constant}) \times \text{CF}] + 200 = \text{Lexile text measure} \qquad \text{Equation (1)}$$

was developed to convert logit difficulties to Lexile calibrations. The values of the conversion factor (CF) and the constant were determined by substituting in the low-end anchor point and then solving the system of equations.

The Lexile Scale ranges from below zero to above 2000L. There is not an explicit bottom or top to the scale, but rather two anchor points on the scale (described above) that describe different levels of reading comprehension. The Lexile Map, a graphic representation of the Lexile Scale from 200L to 1700L, provides a context for understanding reading comprehension.

## **Validity of The Lexile Framework for Reading**

Validity refers to the “degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). In other words, does the test measure what it is supposed to measure. For the Lexile Framework, which measures a skill, the most important aspect of validity that should be examined is construct validity. The validity of the Lexile Framework can be evaluated by examining how well Lexile measures relate to other measures of reading comprehension and text difficulty.

*Lexile Framework and other Measures of Reading Comprehension.* Table 1 presents the results from studies where students were administered a Lexile assessment and another assessment of reading comprehension. There is a strong relationship between reading comprehension ability as measured by the Lexile Framework and reading comprehension ability as measured by other assessments.

Table 1. Results from linking studies conducted with The Lexile Framework for Reading.

Standardized Test	Grades in Study	N	Correlation Between Test Score and Lexile Measure
Stanford Achievement Tests (Ninth Edition)	4, 6, 8, 10	1,167	0.92
Stanford Diagnostic Reading Test (Version 4)	4, 6, 8, 10	1,169	0.91
North Carolina End-of-Grade Test of Reading Comprehension (NCEOG)	3, 4, 5, 8	956	0.90
TerraNova Assessment Series (CTBS/5)	2, 4, 6, 8	2,713	0.92
Texas Assessment of Academic Skills (TAAS)	3 through 8	3,623	0.73 to 0.78*
Gates-MacGinitie Reading Test	2, 4, 6, 8, 10	4,644	0.90
Metropolitan Achievement Test (Eighth Edition)	2, 4, 6, 8, 10	2,382	0.93
Texas Assessment of Knowledge and Skills (TAKS)	3, 5, 8	1,960	0.60 to 0.73*
The Iowa Tests (Iowa Tests of Basic Skills and Iowa Tests of Educational Development)	3, 5, 7, 9, and 11	4,666	0.88
Stanford Achievement Test (Tenth Edition)	2, 4, 6, 8, and 10	3,064	0.93
Oregon Reading/Literature Knowledge and Skills Test	3, 5, 8, and 10	3,180	0.89
Mississippi Curriculum Test	2, 4, 6, and 8	7,045	0.90

Notes: Results are based on final samples used with each linking study.

\*TAAS and TAKS were not vertically equated; separate linking equations were derived for each grade.

*Lexile Framework and the Difficulty of Basal Readers.* In a study conducted by Stenner, Smith, Horabin, and Smith (1987b) Lexile calibrations were obtained for units in 11 basal series. It was presumed that each basal series was sequenced by difficulty. So, for example, the latter portion of a third-grade reader is presumably more difficult than the first portion of the same book. Likewise, a fourth-grade reader is presumed to be more difficult than a third-grade reader. Observed difficulties for each unit in a basal series were estimated by the rank order of the unit in the series. Thus, the first unit in the first book of the first grade was assigned a rank order of one and the last unit of the eighth-grade reader was assigned the highest rank order number.

Correlations were computed between the rank order and the Lexile calibration of each unit in each series. After correction for range restriction and measurement error, the average disattenuated correlation between the Lexile calibration of text comprehensibility and the rank order of the basal units was 0.995 (see *Table 2*).

*Table 2.* Correlations between Theory-Based Calibrations Produced by the Lexile Equation and Rank Order of Unit in Basal Readers.

Basal Series	Number of Units	$r_{OT}$	$R_{OT}$	$R'_{OT}$
Ginn Rainbow Series (1985)	53	.93	.98	1.00
HBJ Eagle Series (1983)	70	.93	.98	1.00
Scott Foresman Focus Series (1985)	92	.84	.99	1.00
Riverside Reading Series (1986)	67	.87	.97	1.00
Houghton-Mifflin Reading Series (1983)	33	.88	.96	.99
Economy Reading Series (1986)	67	.86	.96	.99
Scott Foresman American Tradition (1987)	88	.85	.97	.99
HBJ Odyssey Series (1986)	38	.79	.97	.99
Holt Basic Reading Series (1986)	54	.87	.96	.98
Houghton-Mifflin Reading Series (1986)	46	.81	.95	.98
Open Court Headway Program (1985)	52	.54	.94	.97
Total/Means	660	.839	.965	.995

$r_{OT}$  = raw correlation between observed difficulties (O) and theory-based calibrations (T).

$R_{OT}$  = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction.

$R'_{OT}$  = correlation between observed difficulties (O) and theory-based calibrations (T) corrected for range restriction and measurement error.

\*Mean correlations are the weighted averages of the respective correlations.

Based on the consistency of the results in *Table 2*, the Lexile theory was able to account for the unit rank ordering of the 11 basal series even with numerous differences in the series—prose selections, developmental range addressed, types of prose introduced (i.e., narrative versus expository), and purported skills and objectives emphasized.

*Lexile Framework and the Difficulty of Reading Test Items.* In a study conducted by Stenner, Smith, Horabin, and Smith (1987a), 1,780 reading comprehension test items appearing on nine nationally-normed tests were analyzed. The study correlated empirical item difficulties provided by the publishers with the Lexile calibrations specified by the computer analysis of the text of each item. The empirical difficulties were obtained in one of three ways. Three of the tests included observed logit difficulties from either a Rasch or three-parameter analysis (e.g., NAEP). For four of the tests, logit difficulties were estimated from item  $p$ -values and raw score means and standard deviations (Poznanski, 1990; Wright, and Linacre, 1994). Two of the tests provided no item parameters, but in each case items were ordered on the test in terms of difficulty (e.g., PIAT). For these two tests, the empirical difficulties were approximated by the difficulty rank order of the items. In those cases where multiple questions were asked about a single passage, empirical item difficulties were averaged to yield a single observed difficulty for the passage.

Once theory-specified calibrations and empirical item difficulties were computed, the two arrays were correlated and plotted separately for each test. The plots were checked for unusual

residual distributions and curvature, and it was discovered that the Lexile equation did not fit poetry items or noncontinuous prose items (e.g., recipes, menus, or shopping lists). This indicated that the universe to which the Lexile equation could be generalized was limited to continuous prose. The poetry and noncontinuous prose items were removed and correlations were recalculated. *Table 3* contains the results of this analysis.

*Table 3.* Correlations between Theory-Based Calibrations Produced by the Lexile Equation and Empirical Item Difficulties.

Test	Number of Questions	Number of Passages	Mean	SD	Range	Min	Max	$r_{OT}$	$R_{OT}$	$R'_{OT}$
SRA	235	46	644	353	1303	33	1336	.95	.97	1.00
CAT-E	418	74	789	258	1339	212	1551	.91	.95	.98
Lexile	262	262	771	463	1910	-304	1606	.93	.95	.97
PIAT	66	66	939	451	1515	242	1757	.93	.94	.97
CAT-C	253	43	744	238	810	314	1124	.83	.93	.96
CTBS	246	50	703	271	1133	173	1306	.74	.92	.95
NAEP	189	70	833	263	1162	169	1331	.65	.92	.94
Battery	26	26	491	560	2186	-702	1484	.88	.84	.87
Mastery	85	85	593	488	2135	-586	1549	.74	.75	.77
Total/ Mean	1780	722	767	343	1441	50	1491	.84	.91	.93

$r_{OT}$  = raw correlation between observed difficulties (*O*) and theory-based calibrations (*T*).

$R_{OT}$  = correlation between observed difficulties (*O*) and theory-based calibrations (*T*) corrected for range restriction.

$R'_{OT}$  = correlation between observed difficulties (*O*) and theory-based calibrations (*T*) corrected for range restriction and measurement error.

\*Means are computed on Fisher *Z* transformed correlations.

The last three columns in *Table 3* show the raw correlation between observed (*O*) item difficulties and theoretical (*T*) item calibrations, with the correlations corrected for restriction in range and measurement error. The Fisher *Z* mean of the raw correlations ( $r_{OT}$ ) is 0.84. When corrections are made for range restriction and measurement error, the Fisher *Z* mean disattenuated correlation between theory-based calibration and empirical difficulty in an unrestricted group of reading comprehension items ( $R'_{OT}$ ) is 0.93.

These results show that most attempts to measure reading comprehension, no matter what the item form, type of skill objectives assessed, or response requirement used, measure a common comprehension factor specified by the Lexile theory.

### Text Measure Error Associated with the Lexile Framework

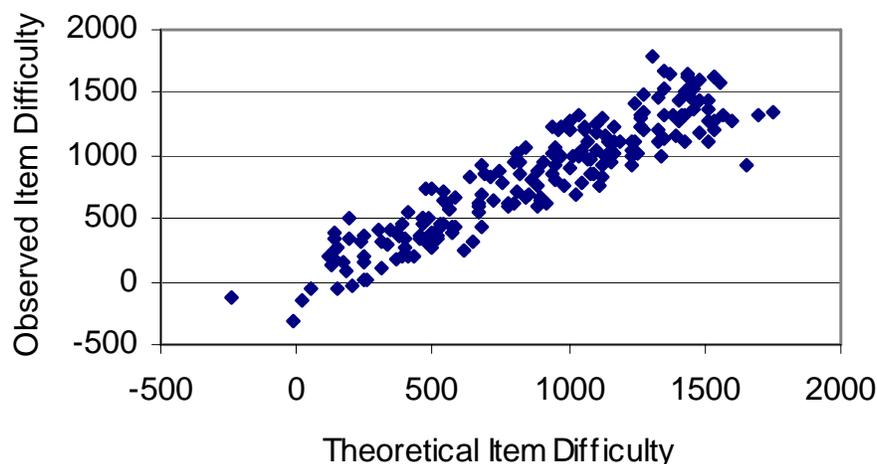
To determine a Lexile measure for a text, the standard procedure is to process the entire text. All pages in the work are concatenated into an electronic file that is processed by a software package called the Lexile Analyzer (developed by MetaMetrics, Inc.). The analyzer “slices” the text file into as many 125-word passages as possible, analyzes the set of slices, and then calibrates each slice in terms of the logit metric. That set of calibrations is then processed to

determine the Lexile measure corresponding to a 75% comprehension rate. The analyzer uses the slice calibrations as test item calibrations and then solves for the measure corresponding to a raw score of 75% (e.g., 30 out of 40 correct, as if the slices were test items). The Lexile Analyzer automates this process, but what “certainty” can be attached to each text measure?

Using the bootstrap procedure to examine error due to the text samples, the above analysis could be repeated. The result would be an identical text measure to the first because there is no sampling error when a complete text is calibrated.

There is, however, another source of error that increases the uncertainty about where a text is located on the Lexile Map. The Lexile Theory is imperfect in its calibration of the difficulty of individual text slices. To examine this source of error, 200 items that had been previously calibrated and shown to fit the model were administered to 3,026 students in Grades 2 through 12 in a large urban school district. For each item the observed item difficulty calibrated from the Rasch model was compared with the theoretical item difficulty calibrated from the regression equation used to calibrate texts. A scatterplot of the data is presented in *Figure 1*.

*Figure 1.* Scatterplot between observed item difficulty and theoretical item difficulty.



The correlation between the observed and the theoretical calibrations for the 200 items was 0.92 and the root mean square error was 178L. Therefore, for an individual slice of text the measurement error is 178L.

The standard error of measurement associated with a text is a function of the error associated with one slice of text (178L) and the number of slices that are calibrated from a text. Very short books have larger uncertainties than longer books. A book with only four slices would have an uncertainty of 89L whereas a longer book such as *War and Peace* (4,082 slices of text) would only have an uncertainty of 3L (*Table 4*).

**Table 4.** Standard errors for selected values of the length of the text.

Title	Number of Slices	Text Measure	Standard Error of Text
<i>The Stories Julian Tells</i>	46	520	26
<i>Bunnicula</i>	102	710	18
<i>The Pizza Mystery</i>	137	620	15
<i>Meditations of First Philosophy</i>	206	1720	12
<i>Metaphysics of Morals</i>	209	1620	12
<i>Adventures of Pinocchio</i>	294	780	10
<i>Red Badge of Courage</i>	348	900	10
<i>Scarlet Letter</i>	597	1420	7
<i>Pride and Prejudice</i>	904	1100	6
<i>Decameron</i>	2431	1510	4
<i>War and Peace</i>	4082	1200	3

A typical Grade 3 reading test has approximately 2,000 words in the passages. To calibrate this text, it would be sliced into 16 125-word passages. The error associated with this text measure would be 45L. A typical Grade 7 reading test has approximately 3,000 words in the passages and the error associated with the text measure would be 36L. A typical Grade 10 reading test has approximately 4,000 words in the passages and the error associated with the text measure would be 30L.

The Lexile Titles Database ([www.Lexile.com](http://www.Lexile.com)) contains information about each book analyzed: author, Lexile measure and Lexile Code, awards, ISBN, and developmental level as determined by the publisher. Information concerning the length of a book and the extent of illustrations—factors that affect a reader’s perception of the difficulty of a book—can be obtained from MetaMetrics.

### **Lexile Item Bank**

The Lexile Item Bank contains over 10,000 items that have been developed between 1986 and 2000 for research purposes with the Lexile Framework.

*Passage Selection.* Passages selected for use are selected from “real world” reading materials that students may encounter both in and out of the classroom. Sources include textbooks, literature, and periodicals from a variety of interest areas and material written by authors of different backgrounds. The following criteria are used to select passages:

- the passage must develop one main idea or contain one complete piece of information;
- understanding of the passage is independent of the information that comes before or after the passage in the source text; and
- understanding of the passage is independent of prior knowledge not contained in the passage.

With the aid of a computer program, item writers examine blocks of text (minimum of three sentences) that are calibrated to be within 100L of the source text. From these blocks of text item writers are asked to select four to five that could be developed as items. If it is necessary to shorten or lengthen the passage in order to meet the criteria for passage selection, the item writer can immediately recalibrate the text to ensure that it is still targeted within 100L of the complete text (source targeting).

*Item Format.* The native Lexile item format is embedded completion. The embedded completion format is similar to the fill-in-the-blank format. When properly written, this format directly assesses the reader's ability to draw inferences and establish logical connections between the ideas in the passage (Haladyna, 1994). The reader is presented with a passage of approximately 30 to 150 words in length. The passages are shorter for beginning readers and longer for more advanced readers. The passage is then response illustrated (a statement is added at the end of the passage with a missing word or phrase followed by four options). From the four presented options, the reader is asked to select the "best" option that completes the statement. With this format, all options are semantically and syntactically appropriate completions of the sentence, but one option is unambiguously the "best" option when considered in the context of the passage.

The statement portion of the embedded completion item can assess a variety of skills related to reading comprehension: paraphrase information in the passage, draw a logical conclusion based on the information in the passage, make an inference, identify a supporting detail, or make a generalization based on the information in the passage. The statement is written to ensure that by reading and comprehending the passage the reader is able to select the correct option. When the embedded completion statement is read by itself, each of the four options is plausible.

*Item Writer Training.* Item writers are classroom teachers and other educators who have had experience with the everyday reading ability of students at various levels. The use of individuals with these types of experiences helped to ensure that the items are valid measures of reading comprehension. Item writers are provided with training materials concerning the embedded completion item format and guidelines for selecting passages, developing statements, and selecting options. The item writing materials also contain incorrect items that illustrate the criteria used to evaluate items and corrections based on those criteria. The final phase of item writer training is a short practice session with three items.

Item writers are provided vocabulary lists to use during statement and option development. The vocabulary lists were compiled from spelling books one grade level below the level the item would typically be used with. The rationale was that these words should be part of a reader's "working" vocabulary since they had been learned the previous year.

Item writers are also given extensive training related to "sensitivity" issues. Part of the item writing materials address these issues and identify areas to avoid when selecting passages and developing items. The following areas are covered: violence and crime, depressing situations/death, offensive language, drugs/alcohol/tobacco, sex/attraction, race/ethnicity, class, gender, religion, supernatural/magic, parent/family, politics, animals/environment, and brand names/junk food. These materials were developed based on material published by McGraw-Hill (*Guidelines for Bias-Free Publishing*). This publication discusses the equal treatment of the sexes, fair representation of minority groups, and the fair representation of disabled individuals.

*Item Review.* All items are subjected to a two-stage review process. First, items are reviewed and edited by an editor according to the 19 criteria identified in the item writing materials and for sensitivity issues. Approximately 25% of the items developed are deleted for various reasons. Where possible items were edited and maintained in the item bank.

Items are then reviewed and edited by a group of specialists that represent various perspectives—test developers, editors, and curriculum specialists. These individuals examine each item for sensitivity issues and for the quality of the response options. During the second stage of the item review process, items are either “approved as presented,” “approved with edits,” or “deleted.” Approximately 10% of the items written are “approved with edits” or “deleted” at this stage. When necessary, item writers receive additional on-going feedback and training.

*Item Analyses.* As part of the linking studies and research studies conducted by MetaMetrics, items in the Lexile Item Bank are evaluated in terms of difficulty (relationship between logit [observed Lexile measure] and theoretical Lexile measure), internal consistency (point-biserial correlation), and bias (ethnicity and gender where possible). Where necessary, items are deleted from the item bank or revised and recalibrated.

During the spring of 1999, 8 levels of a Lexile assessment were administered in a large urban school district to students in grades 1 through 12. The 8 test levels were administered in grades 1, 2, 3, 4, 5, 6, 7-8, and 9-12 and ranged from 40 to 70 items depending on the grade level. A total of 427 items were administered across the 8 test levels. Each item was answered by at least 9,000 students (the number of students per level ranged from 9,286 in grade 2 to 19,056 in grades 9-12). The item responses were submitted to a Winsteps IRT analysis. The resulting item difficulties (in logits) were assigned Lexile measures by multiplying by 180 and anchoring each set of items to the mean theoretical difficulty of the items on the form.

# The AIMS DPA/AIMS HS-Lexile Framework Linking Process

## Description of the Assessments

*Arizona's Instrument to Measure Standards (AIMS)*. AIMS is a “standards-based assessment of Arizona’s Academic Standards in writing, reading, and mathematics” (ADE, 2007). Arizona’s Language Arts Standards were adopted in 1996 and refined and articulated in 2002 and 2003. The standards describe reading as “a complex skill that involves learning language and using it effectively in the active process of constructing meaning embedded in text. It requires students to fluently decode the words on a page, understand the vocabulary of the writer, and use strategies to build comprehension of the text” (ADE, 2003, p.viii). The reading standard consists of three strands:

1. Reading Process – print concepts, phonics, vocabulary, and comprehension,
2. Comprehending Literary Text – elements of literature and historical and cultural aspects of literature, and
3. Comprehending Information Text – expository text, functional text, and persuasive text.

AIMS DPA (Arizona’s Instrument to Measure Standards Dual Purpose Assessment) is both a criterion-referenced and a norm-referenced assessment (CTB/McGraw-Hill, 2007). It is criterion-referenced in that it consists of items written by Arizona teachers that are aligned with Arizona’s Academic Standards. It is norm-referenced in that it includes embedded *TerraNova* items (items identified by Arizona educators as being aligned with the Arizona curriculum). The test assesses reading, language, and mathematics using multiple-choice items. The writing component of the assessment consists of a single prompt essay which is scored using a six-trait analytic rubric. AIMS DPA is administered to all students in grades 3 through 8. The Grade 3 reading test consists of 42 criterion-referenced items and 12 *TerraNova* norm-referenced items for a total of 54 items (Strand 1: 44%, Strand 2: 22%, and Strand 3: 33%). The Grades 4 through 8 reading tests consist of 39 criterion-referenced items and 15 *TerraNova* norm-referenced items for a total of 54 items (Strand 1: 17-23%, Strand 2: 31-33%, and Strand 3: 46-50%). The AIMS DPA is calibrated using the Rasch IRT scale and reading scale scores range from 200 to 800 across Grades 3 through 8.

AIMS HS (Arizona’s Instrument to Measure Standards High School) is a criterion-referenced assessment that is administered to students beginning the spring of their second year in high school (CTB/McGraw-Hill, 2007). The test assesses reading and mathematics using multiple-choice items. The writing component of the assessment consists of a single prompt essay which is scored using a six-trait analytic rubric. The AIMS HS is a high school competency test where passing scores are required to earn a diploma for students who graduated beginning in Spring 2006. Students in Grade 10 have five opportunities to pass the test prior to graduation. The high school reading test consists of 54 criterion-referenced items (Strand 1: 15%, Strand 2: 33%, and Strand 3: 52%). The AIMS HS is calibrated using the Rasch IRT scale and reading scale scores range from 500 to 900.

In 2007, Arizona offered three levels of accommodations: universal, standard, and alternate. Students may have been given a test administration accommodation based on his or her Individualized Education Plan (IEP). Alternate accommodations included provisions that “involve substantial changes on what a student is expected to learn and/or in the way that learning is demonstrated” (ADE, 2007, p. 4). It was considered an alternate accommodation if

the assessment was administered with an “auditory presentation or interpretation through sign language systems of items on the Reading test” (p. 9). The use of alternate accommodations resulted in scores that are considered invalid for comparison and accountability purposes.

*The Lexile Framework for Reading.* The Lexile Framework is a tool that can help teachers, parents, and students locate challenging reading materials. Text difficulty and reader ability are measured in the same unit—the Lexile. Text difficulty is determined by examining such characteristics as word frequency and sentence length. Items and text are calibrated using the Rasch model. The typical range of the Lexile Scale is from 200 to 1700 Lexiles, although actual Lexile measures can range from below zero to above 2000 Lexiles (see the discussion on page 5 for more information).

Using response-illustrated items, the Lexile Framework measures reading comprehension by focusing on skills readers use when studying written materials sampled from various content areas. These skills include referring to details in the passage, drawing conclusions, and making comparisons and generalizations. Lexile items do not require prior knowledge of ideas outside of the passage, vocabulary taken out of context, or formal logic.

Using items from the Lexile Item Bank, Lexile linking tests were developed for administration at grades 3, 5, 7, and HS. A T-parallel test is designed to be an alternate test form such that it is similar in statistical test specifications. Test specifications describe the number of items and the raw score difficulty (mean percent correct). In this study the 2007 AIMS DPA test forms and the 2007 AIMS HS test form are the original test. Each test consisted of 54 items. The *p*-values associated with the items on each grade level form were converted to logits using the P2D computer program (MetaMetrics, Inc., no date) and then scaled to the normative grade level mean (Spring: Grade 3—520L; Grade 5—745L; Grade 7—925L; and HS—1030L) based on other research studies conducted with the Lexile Framework.

*Evaluation of T-parallel Lexile Linking Tests.* After administration, the Lexile Linking Test booklets were reviewed. Based on this examination, one item on the Grade 7 Lexile Linking Test was removed from further analyses (item 30, point-biserial correlation = 0.09). The descriptive statistics of Lexile Linking Test are presented in *Table 5*. The average percent correct for each test is within the expected range of 75%.

*Table 5.* Descriptive statistics from the development of T-parallel assessments.

Grade	<i>N</i>	Percent Correct Range	Point-Biserial Range
3	1,948	0.78 (0.64 to 0.94)	0.27 to 0.65
5	1,961	0.79 (0.64 to 0.95)	0.29 to 0.61
7	1,921	0.77 (0.62 to 0.93)	0.21 to 0.60
HS	1,755	0.69 (0.30 to 0.91)	0.10 to 0.59

For the Lexile Linking Test, raw scores were converted to Lexile measures using an anchored Winsteps analysis for each test form based on the theoretical difficulties of the items on the test.

## Study Design

A single-group design was chosen for this study (Kolen and Brennan, 1995). This design is most useful “when (1) administering two forms to examinees is operationally possible, (2) differential order effects are not expected to occur, and (3) it is difficult to obtain participation of a sufficient number of examinees in an equating study that uses the random groups design” (pp. 16–17). The AIMS HS tests were administered February 27-28, 2007 and the AIMS DPA tests were administered April 9-20, 2007. The AIMS HS/Lexile Linking Tests were administered March 12-16, 2007 and the AIMS DPA/Lexile Linking Tests were administered April 23 to May 4, 2007.

Counterbalancing the order of administration of the tests is one method that is used to minimize differential order effects (Kolen and Brennan, 2004). Two typical causes of order effects are fatigue and test familiarity. In this study counterbalancing was not employed because it was not desirable to administer a unique test to some students prior to the “high stakes” test. The fatigue factor was minimized by having the two tests administered approximately 2 weeks apart. The practice factor was not an issue in this study because the two tests consisted of items employing different formats: AIMS consisted of extended passages followed by multiple stand-alone, curriculum-based items, while the Lexile Linking Test consisted of short paragraph-length passages followed by one embedded-completion statement assessing general reading ability.

## Description of the Sample

The original sample of students for the study was drawn by the Arizona Department of Education. The school districts were located across the state and included both rural and urban districts (Grades 3, 5, and 7: 20 districts; High School: 4 districts). The sample consisted of 43 schools (Grade 3: 24 schools; Grade 5: 24 schools; Grade 7: 18 schools; and High School: 4 schools). *Table 6* presents the samples at each grade level and the proportion of students with complete data (both an Arizona AIMS scale score and a Lexile measure).

*Table 6.* Number of students sampled and the number of students in matched sample.

Grade	Arizona Reading Test Number Sampled	Lexile Linking Test Number Sampled	Matched Sample
3	2,184	1,988	1,935
5	2,191	1,998	1,951
7	2,226	1,943	1,887
High School	2,217	1,806	1,761
<b>Total</b>	<b>8,818</b>	<b>7,735 (87.7%)</b>	<b>7,534 (85.4%)</b>

*Table 7* presents the demographics characteristics of all students in Grades 3, 5, 7, and High School (cohort09) who were administered the AIMS in spring 2007.

*Table 7.* Demographic characteristics of all Arizona students, spring 2007, Grades 3, 5, 7, and High school (cohort09).

Student Characteristic	Grade 3 (N = 81,431)	Grade 5 (N = 81,616)	Grade 7 (N = 80,820)	High School (N = 74,324)
<i>Gender</i>				
Female	48.7%	49.0%	48.9%	49.0%
Male	51.1%	50.*%	50.9%	50.9%
<i>Ethnicity</i>				
American Indian	5.1%	5.2%	5.5%	6.1%
Asian	2.8%	2.7%	2.6%	2.7%
Black/not Hispanic	5.6%	5.5%	5.5%	5.5%
Hispanic	43.4%	41.9%	40.9%	36.3%
White/not Hispanic	43.1%	44.6%	45.4%	49.4%
<i>Language</i>				
English	69.9%	70.7%	72.3%	76.6%
Spanish	26.3%	25.7%	24.1%	18.9%
Navajo	1.0%	1.1%	1.1%	1.2%
Other Language	2.4%	2.2%	2.2%	2.1%
Unknown Language	0.4%	0.3%	0.4%	1.1%
<i>Special Status</i>				
Limited English Proficient	16.4%	11.4%	9.5%	4.8%
Fluent English Proficient	5.0%	5.3%	2.7%	1.5%
Disabled	11.1%	11.9%	10.7%	8.7%

The matched sample used in this study consisted of students who had both AIMS Reading scale scores and Lexile measures. Additional students were removed from the sample based on the infit and outfit parameters from the Winsteps analysis with the Lexile Linking Test results. All students and items were submitted to a Winsteps analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.003. A total of 1,284 students in Grades 3, 5, 7, and High School were removed from further analyses. The sample was reduced by 14.6%. *Table 8* presents the demographics characteristics of the matched sample of students in Grades 3, 5, 7, and High School (cohort09) used in the linking study.

**Table 8.** Demographic characteristics of all Arizona students, matched sample spring 2007, Grades 3, 5, 7, and High school (cohort09).

Student Characteristic	Grade 3 (N = 1,935)	Grade 5 (N = 1,951)	Grade 7 (N = 1,887)	High School (N = 1,761)
<i>Gender</i>				
Female	48.0%	47.2%	47.0%	50.0%
Male	51.9%	52.7%	52.9%	50.0%
<i>Ethnicity</i>				
American Indian	7.8%	8.8%	10.2%	4.0%
Asian	2.7%	2.2%	1.4%	2.2%
Black/not Hispanic	4.3%	3.8%	4.1%	5.2%
Hispanic	39.3%	37.9%	38.6%	55.0%
White/not Hispanic	45.8%	47.3%	45.5%	33.7%
<i>Language</i>				
English	77.4%	77.0%	78.1%	71.5%
Spanish	19.0%	18.7%	19.1%	27.2%
Navajo	2.2%	2.7%	1.6%	0.1%
Other Language	1.3%	1.3%	1.1%	1.0%
Unknown Language	0.1%	0.3%	0.2%	0.2%
<i>Special Status</i>				
Limited English Proficient	14.6%	9.3%	6.5%	3.8%
Fluent English Proficient	4.0%	4.2%	2.3%	1.4%
Disabled	9.1%	9.9%	9.5%	7.0%

*Table 9* presents the descriptive statistics for each assessment for the matched sample. *Figures 2* through *4* show the relationship between AIMS Reading scale scores and Lexile measures from the Lexile Linking Tests in Grades 3, 5, and 7. *Figure 5* shows the relationship between AIM DPA Reading scale score and Lexile measure from the Lexile Linking Test. *Figure 6* shows the relationship between AIM HS Reading scale score and Lexile measure from the Lexile Linking Test.

**Table 9.** Descriptive statistics for the AIMS DPA Reading scale scores and the Lexile Linking Test results, matched sample, Grades 3, 5, 7, and High School (cohort09).

Grade	N	AIMS Reading Scale Scores Mean (SD)	Lexile Linking Test Mean (SD)	r
3	1,935	457.81 (49.71)	536.90 (276.12)	0.827
5	1,951	491.78 (44.53)	801.15 (281.83)	0.840
7	1,887	513.49 (50.28)	1003.00 (241.34)	0.841
High School	1,761	694.96 (45.53)	1123.00 (250.87)	0.790

Figure 2. Scatter plot of Grade 3 AIMS DPA Reading scale scores and Lexile measures, matched sample ( $N = 1,935$ ).

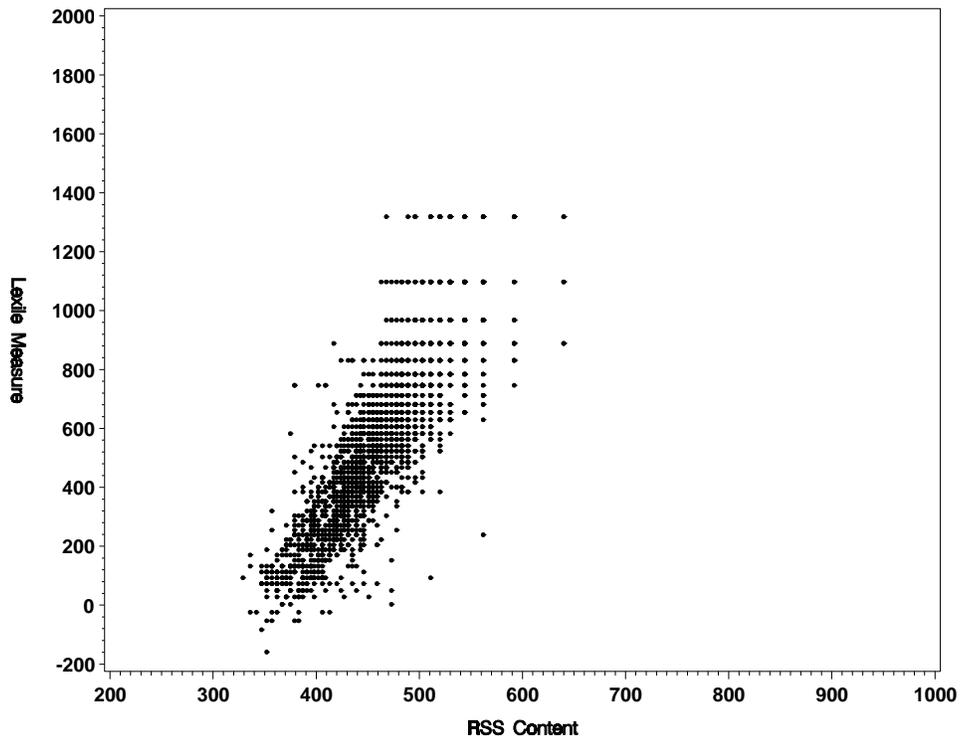


Figure 3. Scatter plot of Grade 5 AIMS DPA Reading scale scores and Lexile measures, matched sample ( $N = 1,951$ ).

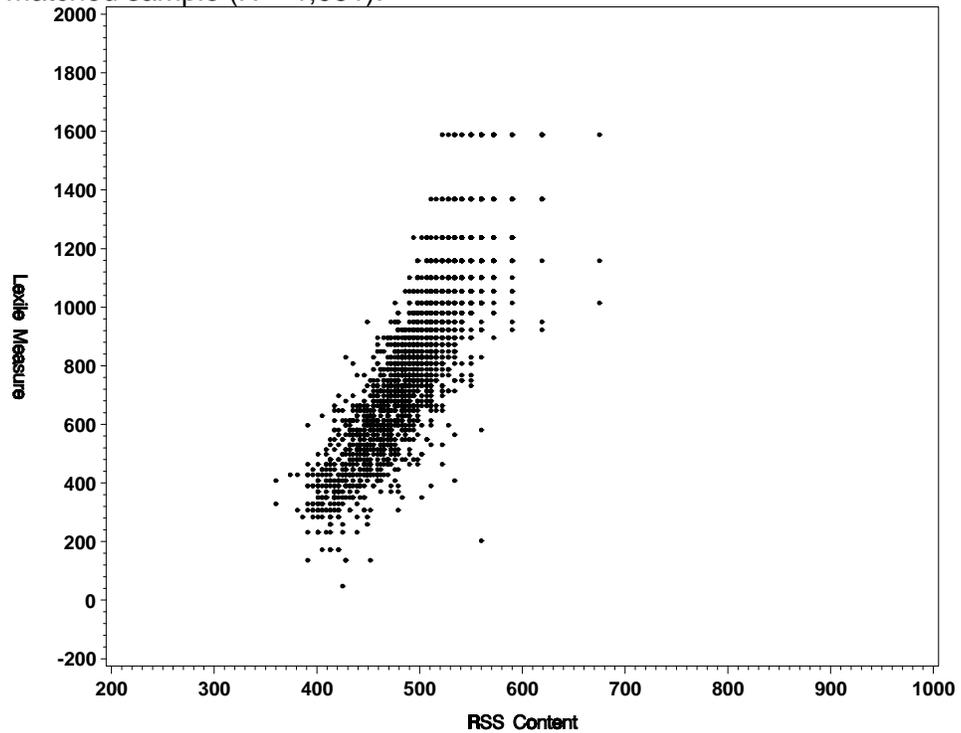


Figure 4. Scatter plot of Grade 7 AIMS DPA Reading scale scores and Lexile measures, matched sample ( $N = 1,887$ ).

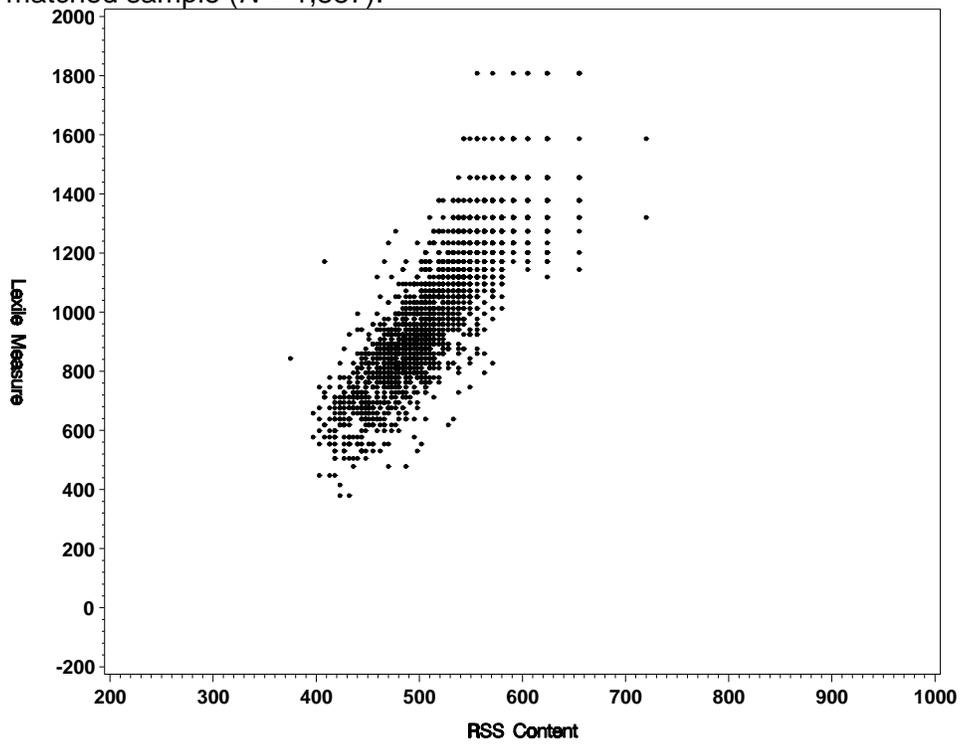


Figure 5. Scatter plot of Grades, 3, 5, and 7 AIMS DPA Reading scale scores and Lexile measures, matched sample ( $N = 5,773$ ).

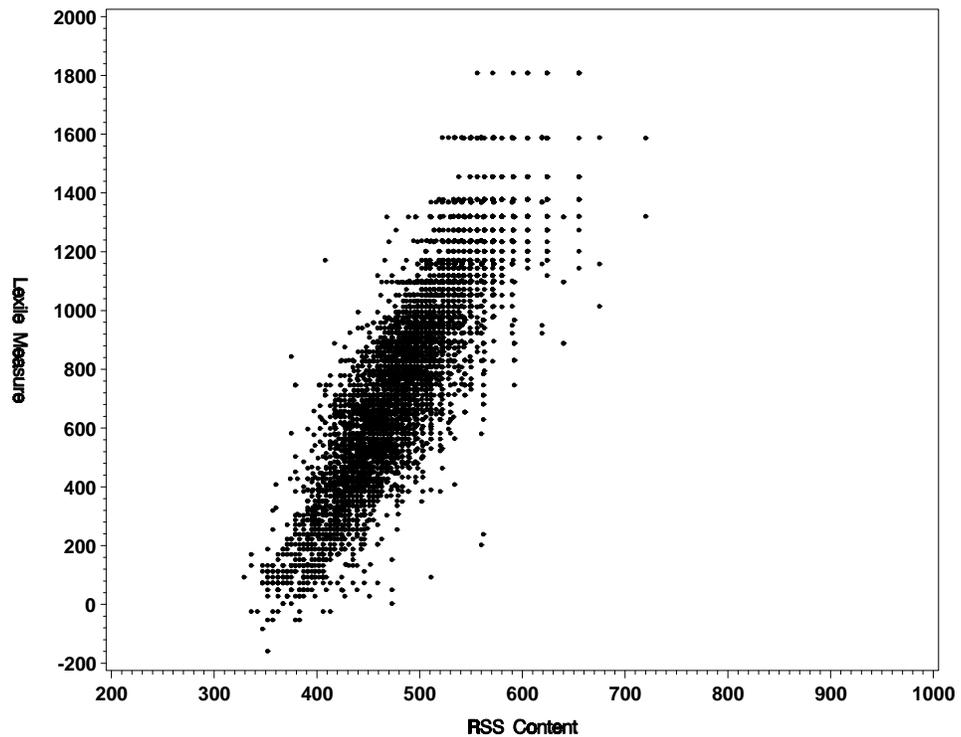
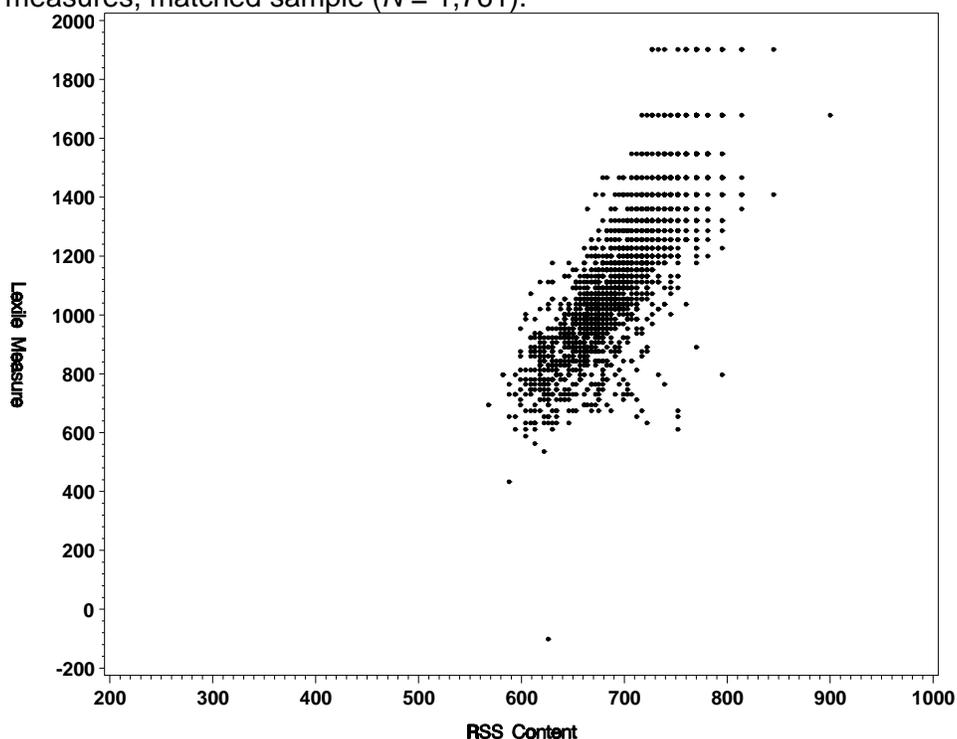


Figure 6. Scatter plot of High School (cohort09) AIMS HS Reading scale scores and Lexile measures, matched sample (N = 1,761).



For the final sample of students used in the study, students with the following score patterns were removed:

- 0% correct (raw score of 0) on the AIMS or the Lexile Linking Test,
- 100% correct on the AIMS or the Lexile Linking Test,
- Scoring at or below chance on the AIMS.

A total of 242 students (2.44% of the matched sample) were removed.

To account for individual differences in motivation when responding to the two assessments, the sample set was further trimmed. Test scores from each of the assessments were rank ordered and then converted to percentiles. For each student, the difference in percentiles between the two assessments was examined. The 25-percentile difference was selected to minimize the number of students removed from the sample and maintain the characteristics of the distribution, while at the same time removing students that were obvious outliers on one or both of the assessments. A total of 927 students were removed from further analyses (Grades 3, 5, and 7: 667 students; High School (cohort09): 260 students). The sample was reduced by 12.3%.

Finally, a total of 599 students were removed from further analyses (Grades 3, 5, and 7: 523 students; High School (cohort09): 76 students) because of accommodations used during the test administrations (classified on student answer sheet) or because of English Language Learner status. Students classified as Limited English Proficient (students whose English proficiency is still developing) were excluded from further analyses; while, students classified as Fluent English Proficient (students that score at the proficient level of the state mandated

English language proficiency assessment) were included in further analyses. The sample was reduced by 7.95%. The resulting samples consisted of 4,249 students in Grades 3 through 7 (71.7%) and 1,350 students in High School (74.8%). The sample statistics did not change because of the exclusion of these two groups of students. *Table 10* presents the demographic characteristics and *Table 11* presents the descriptive statistics of the final (trimmed) sample by grade.

*Table 10.* Demographic characteristics of all Arizona students, final (trimmed) sample spring 2007, grades 3, 5, 7, and High school (cohort09).

Student Characteristic	Grade 3 (N = 1,317)	Grade 5 (N = 1,455)	Grade 7 (N = 1,477)	High School (N = 1,350)
<i>Gender</i>				
Female	49.4%	48.0%	47.6%	51.6%
Male	50.5%	51.8%	52.3%	48.4%
<i>Ethnicity</i>				
American Indian	5.5%	6.9%	10.5%	4.4%
Asian	2.6%	2.0%	1.1%	2.2%
Black/not Hispanic	5.5%	4.5%	4.3%	4.6%
Hispanic	35.6%	36.2%	35.5%	54.0%
White/not Hispanic	50.9%	50.4%	48.5%	34.8%
<i>Language</i>				
English	85.3%	81.9%	82.3%	73.3%
Spanish	12.9%	15.3%	15.0%	25.6%
Navajo	0.5%	1.2%	1.8%	0.1%
Other Language	1.1%	1.4%	0.8%	0.7%
Unknown Language	0.1%	0.3%	0.2%	0.3%
<i>Special Status</i>				
Fluent English Proficient	4.8%	4.7%	2.4%	1.5%
Disabled	8.7%	9.6%	7.8%	6.1%

*Table 11.* Descriptive statistics for the AIMS DPA Reading scale scores and the Lexile Linking Test results, final (trimmed) sample, Grades 3, 5, 7, and High School (cohort09).

Grade	N	AIMS Reading Scale Scores Mean (SD)	Lexile Linking Test Mean (SD)	r
3	1,317	463.53 (45.66)	587.91 (239.31)	0.889
5	1,455	495.58 (40.26)	822.12 (232.86)	0.899
7	1,477	519.10 (45.26)	990.35 (214.32)	0.890
High School	1,350	698.19 (42.97)	1131.00 (218.46)	0.887

Figures 7 through 9 show the relationship between AIMS Reading scale scores and Lexile measures from the Lexile Linking Tests in Grades 3, 5, and 7 for the final (trimmed) sample. Figure 10 shows the relationship between AIM DPA Reading scale score and Lexile measure from the Lexile Linking Test for the final (trimmed) sample. Figure 11 shows the relationship between AIM HS Reading scale score and Lexile measure from the Lexile Linking Test for the final (trimmed) sample.

Figure 7. Scatter plot of Grade 3 AIMS DPA Reading scale scores and Lexile measures, final (trimmed) sample ( $N = 1,317$ ).

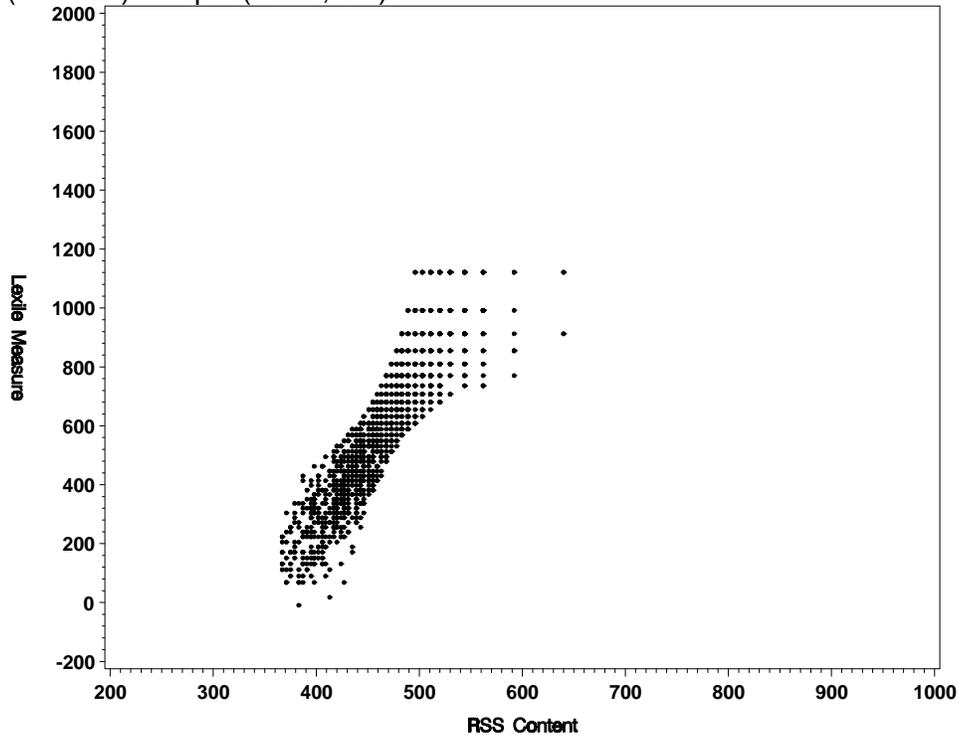


Figure 8. Scatter plot of Grade 5 AIMS DPA Reading scale scores and Lexile measures, final (trimmed) sample ( $N = 1,455$ ).

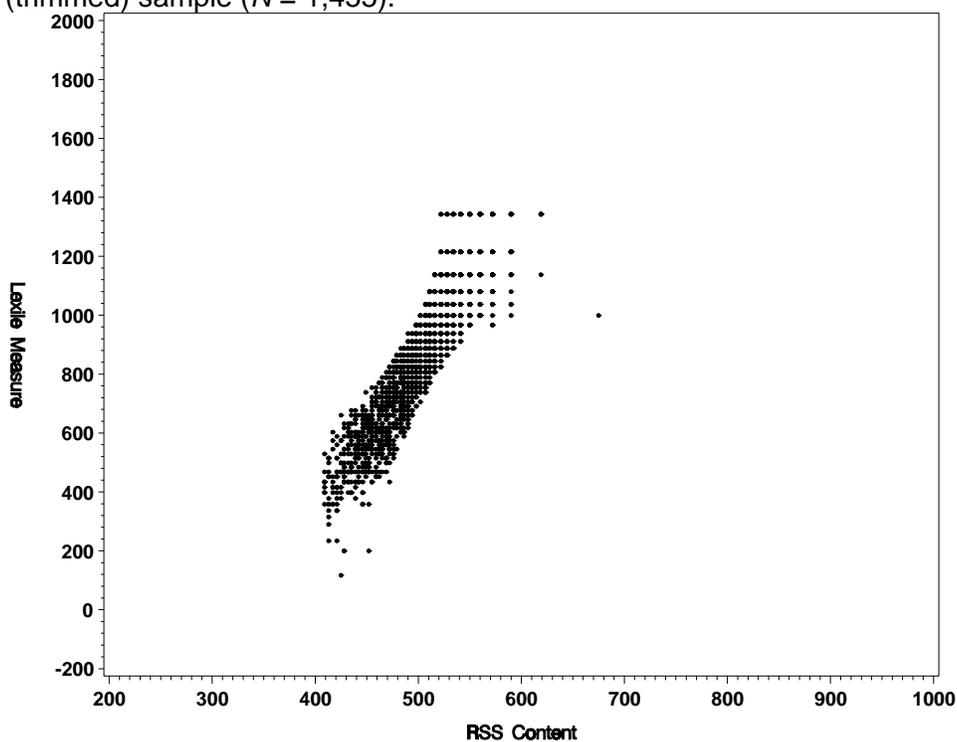


Figure 9. Scatter plot of Grade 7 AIMS DPA Reading scale scores and Lexile measures, final (trimmed) sample ( $N = 1,477$ ).

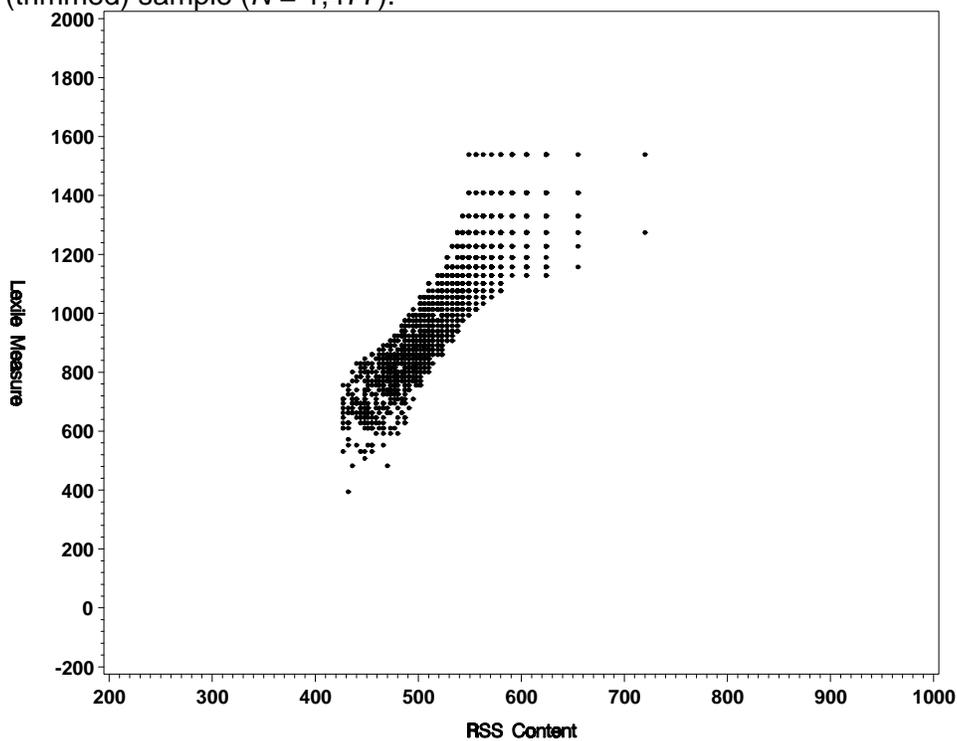


Figure 10. Scatter plot of Grades 3, 5, and 7 AIMS DPA Reading scale scores and Lexile measures, final (trimmed) sample ( $N = 4,249$ ).

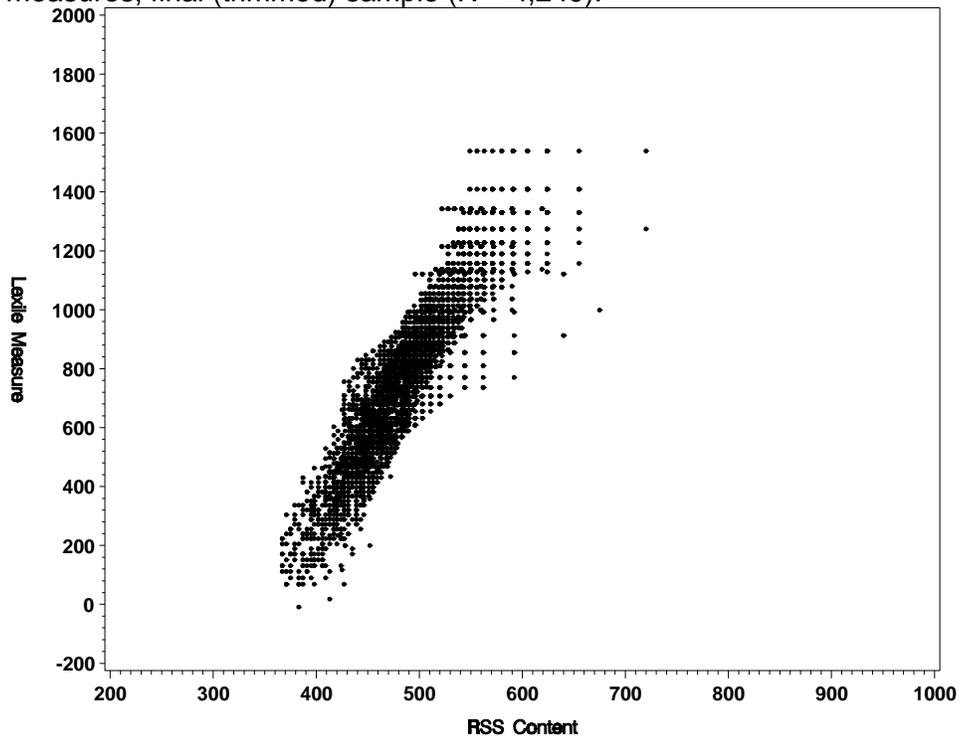
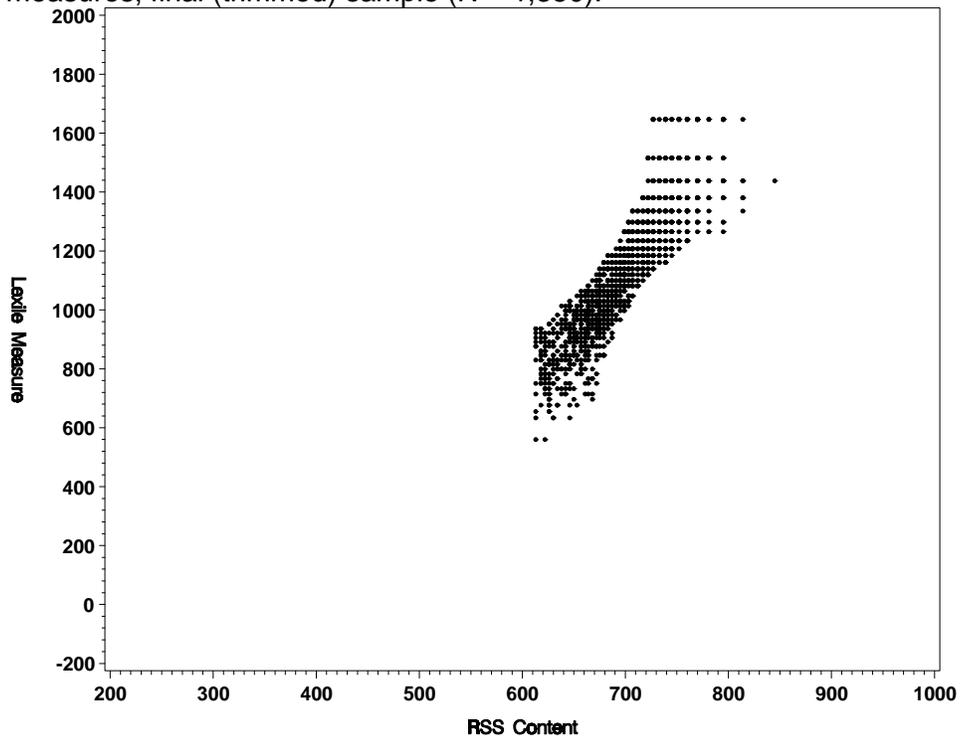


Figure 11. Scatter plot of High School (cohort09) AIMS HS Reading scale scores and Lexile measures, final (trimmed) sample ( $N = 1,350$ ).



## Linking the Arizona AIMS Reading Scale Scores (AIMS DPA and AIMS HS) with the Lexile Scale

Linking in general means “putting the scores from two or more tests on the same scale” (National Research Council, 1999, p.15). MetaMetrics, Inc. and the Arizona Department of Education conducted this linking study for the purpose of matching students with books—to predict the books a student should be matched with for successful reading experiences, given their performance on the AIMS DPA or the AIMS HS.

*Evaluation of linkage assumptions.* Factors that affect the linkage between two assessments include the domain to be assessed, the definition of the framework for assessment, the test specifications, and the items sampled.

Based upon the correlations between Arizona AIMS DPA and AIMS HS scale scores and the Lexile measures presented in *Tables 9 and 11*, it can be concluded that the two assessments measure very similar constructs (reading comprehension). The correlations between the two assessments are within the typical range of the alternate-form reliability coefficients; therefore, the Lexile Linking Test can be considered a T-parallel form of Arizona AIMS DPA/AIMS HS. By using alternate-form reliability coefficients as a comparison, similar sources of variation are accounted for (differences in testing occasions and items).

In addition, the tests were constructed to have the same number of items and the same level of difficulty.

*Linking Analyses.* Two score scales (e.g., the Lexile Scale and the Arizona Reading Scale) can be linked using linear equating when the underlying item response models used to develop assessments are the same or different. The linear equating method is most appropriate when (1) sample sizes are small; (2) test forms have similar difficulties; and (3) simplicity in conversion tables or equations, in conducting analyses, and in describing procedures are desired (Kolen and Brennan, 1995).

In linear equating, a transformation is chosen such that scores on two tests are considered to be equated if they correspond to the same number of standard deviations above (or below) the mean in some group of examinees (Angoff, 1984, cited in Petersen, Kohen, and Hoover, 1989; Kolen and Brennan, 1995). Given scores  $x$  and  $y$  on tests  $X$  and  $Y$ , the linear relationship is

$$\frac{(x - \mu_x)}{\sigma_x} = \frac{(y - \mu_y)}{\sigma_y} \quad (\text{Equation 2})$$

and the linear transformation  $I_x$  (called the SD line in this report) used to transform scores on test  $X$  to scores on text  $Y$  is

$$x = I_x(y) = \left( \frac{\sigma_x}{\sigma_y} \right) y + \left( \mu_x - \frac{\mu_y \sigma_x}{\sigma_y} \right) \quad (\text{Equation 3})$$

Linear equating using an SD-line approach is preferable to linear regression because the tests are not perfectly correlated. With less than perfectly reliable tests, linear regression is dependent on which way the regression is conducted: predicting scores on test  $X$  from scores

on test Y or predicting scores on test Y from scores on test X. The SD line provides the symmetric linking function that is desired.

The final linking equation between the Arizona AIMS Reading scale scores and the Lexile Scale can be written as:

$$\text{Lexile measure} = \text{Slope}(\text{Arizona Reading scale score}) + \text{constant}_g \quad (\text{Equation 4})$$

where the slope is the ratio of the standard deviations of the scale scores from the two assessments.

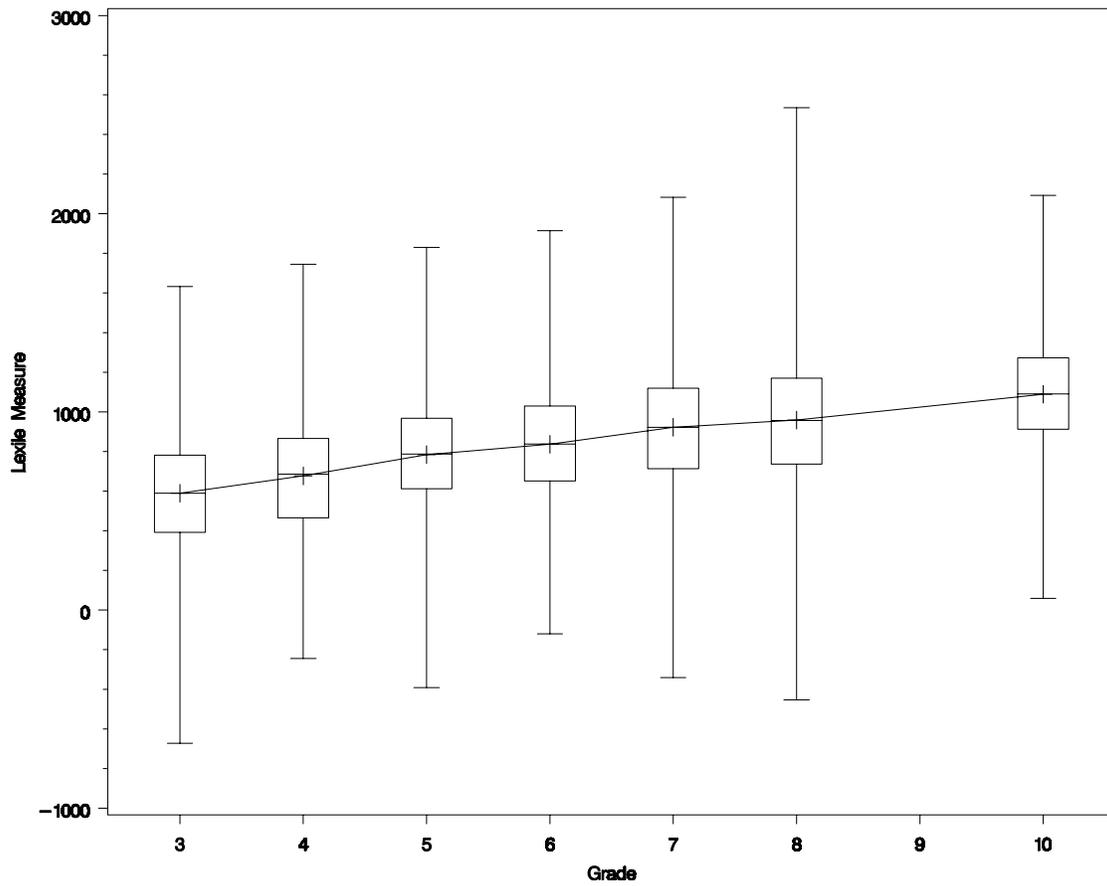
Using the final sample date described in *Table 11*, the linear regression equations relating the Arizona AIMS DPA and AIMS HS Reading scale scores and Lexile measures for all students in the sample by grade are presented in *Table 12*. The intercept in the High School (AIMS HS) linking function was reduced by 65L. Conversion tables were developed in order to express the Arizona Reading scale scores from the AIMS DPA and the AIMS HS in the Lexile metric and were provided to the Arizona Department of Education in electronic format.

*Table 12.* Linear linking equations used to predict Lexile measures from Arizona (AIMS DPA and AIMS HS) Reading scale scores.

Grades (g)	Slope of Linear Linking Equation	Intercept of Linear Linking Equation (constant <sub>g</sub> )
3-8	5.6357250739	-1975.029915
High School	5.0844224233	-2484.134844

For the students administered the AIMS DPA and AIMS HS in spring 2007 ( $N = 559,311$ ), the distribution of predicted Lexile measures across grades/levels is shown in *Figure 12*.

Figure 12. Box-and-whisker plot of predicted Lexile measures from the Arizona AIMS DPA and the AIMS HS.



## Validity of the Arizona AIMS DPA and AIMS HS Link

The Arizona Department of Education established four student performance standards for each grade to describe student performance in terms of what the student has learned and can do based upon the utilization of grade-level appropriate reading materials (Arizona Department of Education, 2005). For each grade level, four standards were established: Exceeds the Standard, Meets the Standard, Approaches the Standard, and Falls Far Below the Standard. Below are the performance level descriptors for Grade 3 Reading.

- *Exceeds the Standard:* Students who score in this level illustrate a superior academic performance as evidenced by achievement that is substantially beyond the goal for all students. Students who perform at this level demonstrate the ability to determine the meaning of words and phrases using context clues, use reading comprehension strategies to draw conclusions and analyze literary elements, and evaluate informational text to determine fact from opinion.
- *Meets the Standard:* Students who score in this level demonstrate a solid academic performance on subject matter as reflected by the reading standard. Students who perform at this level are able to identify character traits, setting, and the sequence of events. In addition to noting the topic sentence in a paragraph, they are able to identify the main idea and supporting details in informational text.
- *Approaches the Standard:* Students who score in this level show partial understanding of the knowledge and application of skills that are fundamental for proficient work. Students who perform at this level show some understanding of decoding skills, using pictures and information from the text to determine the meaning of simple words. They are able to identify images which appeal to the senses and repetition that is utilized in poetry. Some gaps in knowledge and skills are evident and may require additional instruction and remediation in order to achieve a satisfactory level of understanding.
- *Falls Far Below the Standard:* Students who score in this level may have significant gaps and limited knowledge and skills that are necessary to satisfactorily meet the state's reading standard. Students will usually require a considerable amount of additional instruction in order to achieve a satisfactory level of understanding.

*Table 13* shows the Arizona Reading performance standards and the associated Lexile measure for each grade. The performance standards show a monotonically increasing function as grade level increases.

Table 13. Arizona Reading performance standard cut scores in the Lexile metric.

Grade	Meets the Standard		Exceeds the Standard	
	AIMS Reading SS	Lexile measure	AIMS Reading SS	Lexile measure
3	431	455L	516	935L
4	450	560L	536	1045L
5	468	660L	556	1160L
6	478	720L	571	1245L
7	489	780L	587	1335L
8	499	835L	602	1420L
High School	674	945L	773	1445L

### The Lexile Framework and Forecasted Comprehension Rates

A reader with a measure of 600L who is given a text measured at 600L is expected to have a 75-percent comprehension rate. This 75-percent comprehension rate is the basis for selecting text that is targeted to a reader's reading ability, but what exactly does it mean? And what would the comprehension rate be if this same reader were given a text measured at 350L or one at 850L?

The 75-percent comprehension rate for a reader-text pairing can be given an operational meaning by imagining the text is carved into item-sized slices of approximately 125-140 words with a question embedded in each slice. A reader who answers three-fourths of the questions correctly has a 75-percent comprehension rate.

Suppose instead that the text and reader measures are not the same. It is the difference in Lexiles between reader and text that governs comprehension. If the text measure is less than the reader measure, the comprehension rate will exceed 75 percent. If not, it will be less. The question is "By how much?" What is the expected comprehension rate when a 600L reader reads a 350L text?

If all the item-sized slices in the 350L text had the same calibration, the 250L difference between the 600L reader and the 350L text could be determined using the Rasch model equation. This equation describes the relationship between the measure of a student's level of reading comprehension and the calibration of the items. Unfortunately, comprehension rates calculated by this procedure would be biased because the calibrations of the slices in ordinary prose are not all the same. The average difficulty level of the slices *and* their variability both affect the comprehension rate.

Although the exact relationship between comprehension rate and the pattern of slice calibrations is complicated, Equation 5 is an unbiased approximation:

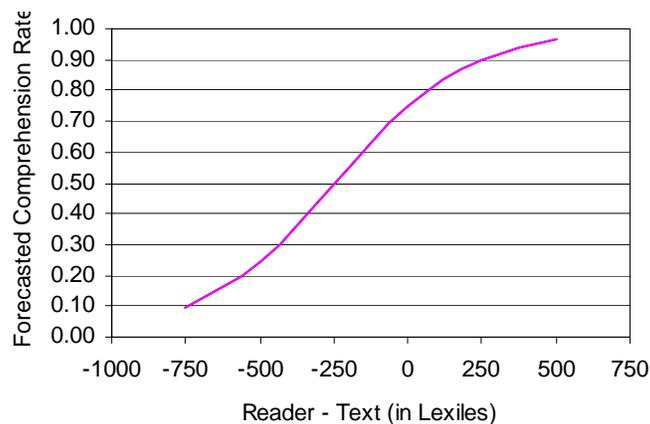
$$\text{Rate} = \frac{e^{\text{ELD}+1.1}}{1 + e^{\text{ELD}+1.1}} \quad (\text{Equation 5})$$

where ELD is the “effective logit difference” given by

$$\text{ELD} = (\text{Reader Lexile measure} - \text{Text Lexile measure}) \div 225. \quad (\text{Equation 6})$$

*Figure 13* shows the general relationship between reader-text discrepancy and forecasted comprehension rate. When the reader measure and the text calibration are the same (difference of 0L) then the forecasted comprehension rate is 75 percent. In the example in the preceding paragraph, the difference between the reader measure of 600L and the text calibration of 350L is 250L. Referring to *Figure 13* and using +250L (reader minus text), the forecasted comprehension rate for this reader-text combination would be 90 percent.

*Figure 13.* Relationship between Reader-Text Discrepancy and Forecasted Comprehension Rate.



*Tables 14* and *15* show comprehension rates calculated for various combinations of reader measures and text calibrations.

**Table 14.** Comprehension rates for the same individual with materials of varying comprehension difficulty.

Person Measure	Text Calibration	Sample Titles	Forecast Comprehension
1000	500	<i>Tornado</i> (Byars)	96%
1000	750	<i>The Martian Chronicles</i> (Bradbury)	90%
1000	1000	<i>Reader's Digest</i>	75%
1000	1250	<i>The Call of the Wild</i> (London)	50%
1000	1500	<i>On the Equality Among Mankind</i> (Rousseau)	25%

**Table 15.** Comprehension rates of different ability persons with the same material.

Person Measure	Calibration for a Grade 10 Biology textbook	Forecast Comprehension Rate
500	1000	25%
750	1000	50%
1000	1000	75%
1250	1000	90%
1500	1000	96%

The subjective experience of 50 percent, 75 percent, and 90 percent comprehension as reported by readers varies greatly. A 1000L reader reading 1000L text (75 percent comprehension) reports confidence and competence. Teachers listening to such a reader report that the reader can sustain the meaning thread of the text and can read with motivation and appropriate emotion and emphasis. In short, such readers appear to comprehend what they are reading. A 1000L reader reading 1250L text (50 percent comprehension) encounters so much unfamiliar vocabulary and difficult syntactic structures that the meaning thread is frequently lost. Such readers report frustration and seldom choose to read independently at this level of comprehension. Finally, a 1000L reader reading 750L text (90 percent comprehension) reports total control of the text, reads with speed, and experiences automaticity during the reading process.

The primary utility of the Lexile Framework is its ability to forecast what happens when readers confront text. With every application by teacher, student, librarian, or parent there is a test of the Framework's accuracy. The Framework makes a point prediction every time a text is chosen for a reader. Anecdotal evidence suggests that the Lexile Framework predicts as

intended. That is not to say that there is an absence of error in forecasted comprehension. There is error in text measures, reader measures, and their difference modeled as forecasted comprehension. However, the error is sufficiently small that the judgments about readers, texts, and comprehension rates are useful.

*Relationship between Linking Error and Forecasted Comprehension Rate.* Using Equation 5 with different combinations of reader measure and text difficulty, the effect of linking error on forecasted comprehension rate can be examined. *Table 16* shows the changes in the forecasted comprehension rate for different combinations of reader and text interactions. When the linking error is small, 5–10L, then the effect on forecasted comprehension rate is a minimal difference (1 to 2 percent) increase or decrease in comprehension.

*Table 16.* Effect of reader-text discrepancy on forecasted comprehension rate.

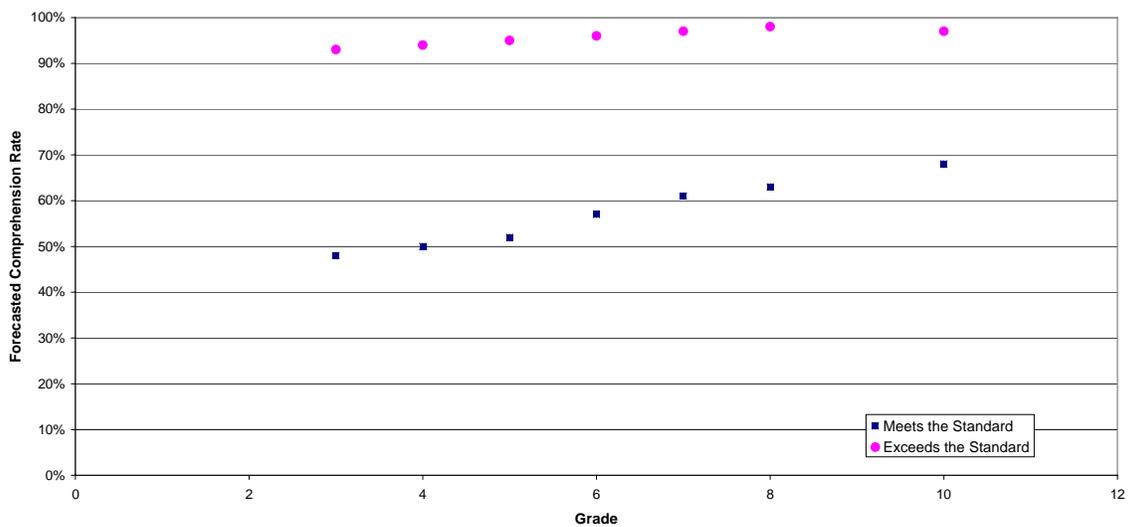
Reader Lexile Measure	Text Lexile Measure	Difference	Forecasted Comprehension Rate
1000L	970L	30L	77.4%
1000L	975L	25L	77.0%
1000L	980L	20L	76.7%
1000L	985L	15L	76.3%
1000L	990L	10L	75.8%
1000L	995L	5L	75.4%
1000L	1000L	0L	75.0%
1000L	1005L	-5L	74.6%
1000L	1010L	-10L	74.2%
1000L	1015L	-15L	73.8%
1000L	1020L	-20L	73.3%
1000L	1025L	-25L	72.9%
1000L	1030L	-30L	72.4%

One way to examine the consistency of the Arizona performance standards is to determine the forecasted comprehension of “grade-level” reading demands. The Lexile Calculator is a utility designed to forecast the reader’s expected comprehension of a specific text (available at [www.Lexile.com](http://www.Lexile.com)). For each grade level, a “book bag” of textbooks was used as the comparison. A “book bag” is a set of the typical books used by a student in a specific grade. While the titles of the specific textbooks may vary from district to district or school to school, the reading demands of the textbooks in the “book bag” are similar within a content area. *Table 17* and *Figure 14* present the forecasted comprehension rates associated with the performance standards set for each grade level.

Table 17. Effect of reader-text discrepancy on forecasted comprehension rate using Arizona Reading performance standards and a grade-appropriate “book bag”.

Grade	Meets the Standard		Exceeds the Standard	
	Lexile measure	Forecasted Comprehension Rate	Lexile measure	Forecasted Comprehension Rate
3	455L	48%	935L	93%
4	560L	50%	1045L	94%
5	660L	52%	1160L	95%
6	720L	57%	1245L	96%
7	780L	61%	1335L	97%
8	835L	63%	1420L	98%
High School	945L	68%	1445L	97%

Figure 14. Forecasted reading comprehension rate for Arizona Reading performance standards.



## Conclusions, Caveats, and Recommendations

Forging a link between scales is a way to add value to one scale without having to administer an additional test. Value can be in the form of any or all of the following:

- increased *interpretability* (e.g., “Based on this test score, what can my child actually read?”),
- increased *diagnostic capability* (e.g., “Based on this test score, what are the student’s weaknesses?”), or
- increased *instructional use* (e.g., “Based on these test scores, I need to modify my instruction to include these skills.”).

The link that has been established between the AIMS DPA and AIMS HS Reading scale scores and the Lexile scale permits readers to be matched with books that provide an appropriate level of challenge while avoiding frustration. The result of this purposeful match may be that students will read more, and, thereby read better. The real power of the Lexile Framework is in examining the growth of readers—wherever the reader may be in the development of his or her reading skills. Readers can be matched with texts that they are forecasted to read with 75 percent comprehension. As a reader grows, he or she can be matched with more demanding texts. And, as the texts become more demanding, then the reader grows.

The development of the link between the AIMS DPA and AIMS HS Reading scale scores and the Lexile scale has been described and evaluated in this study. There are many factors that can affect the linking process. In this study these factors include:

- sample characteristics (e.g., ethnicity), and
- relationship of sample distribution characteristics to the distribution characteristics of the standardization sample.

*Recommendations about reporting Lexile measures for readers.* Lexile measures are reported as a number followed by a capital “L” for “Lexile.” There is no space between the measure and the “L,” and measures of 1,000 or greater are reported without a comma (e.g., 1050L). All Lexile measures should be rounded to the nearest 5L to avoid over interpretation of the measures. As with any test score, uncertainty in the form of measurement error is present.

The measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is accountability (at the student, school, or district level), then actual measures should be reported at all score points. If the purpose is instructional, then the scores should be capped at the lower and upper bounds of measurement error (e.g., at the 25<sup>th</sup> [chance] percentile point and the 95<sup>th</sup> percentile). In an instructional environment where the purpose of the Lexile measure is to appropriately match readers with books, all scores at or below 0L should be reported as “BR” (Beginning Reader); no student should receive a negative Lexile measure. The final reporting ranges are presented in *Table 18*.

Table 18. Reporting range of Lexile measures for instructional purposes.

Grade	Minimum		Maximum	
	Arizona Reading Scale Score	Lexile measure	Arizona Reading Scale Score	Lexile measure
3	200	BR	522 and Above	965L
4	220	BR	539 and Above	1060L
5	240	BR	556 and Above	1160L
6	250	BR	571 and Above	1245L
7	260	BR	587 and Above	1335L
8	270	BR	602 and Above	1420L
High School	500	60L	784 and Above	1500L

*Next Steps.* To utilize the results from this study, Lexile measures need to be incorporated into the Arizona AIMS DPA and AIMS HS results processing and interpretation frameworks. This information can then be used in a variety areas within the educational system—instruction, assessment, communication to name a few.

Within the *instructional area*, suggested book lists can be developed for ranges of readers. Care must be taken to ensure that the books on the lists are also developmentally appropriate for the readers. The Lexile measure is one factor related to comprehension and is a good starting point in the selection process of a book for a specific reader. Other factors such as student developmental level, motivation, and interest; amount of background knowledge possessed by the reader; and characteristics of the text such as illustrations and formatting also need to be considered when matching a book with a reader.

In this era of student-level accountability and high-stakes assessment, differentiated instruction—the attempt “on the part of classroom teachers to meet students where they are in the learning process and move them along as quickly and as far as possible in the context of a mixed-ability classroom” (Tomlinson, 1999)—is a means for all educators to help students succeed. Differentiated instruction promotes high-level and powerful curriculum for all students, but varies the level of teacher support, task complexity, pacing, and avenues to learning based on student readiness, interest, and learning profile. One strategy for managing a differentiated classroom suggested by Tomlinson is the use of multiple texts and supplementary materials.

In a differentiated classroom, teachers must make decisions about a student’s readiness and then design a learning experience based upon their best understanding of the student’s needs and interest. One of the stated purposes of the AIMS DPA and AIMS HS is school improvement. This purpose can be met through the use of the Lexile Framework to match the text environment of the classroom (instructional materials) with the reading levels of the students. The Lexile Framework is an objective tool that can be used to determine a student’s

readiness for a reading experience; the Lexile Framework “targets” text (books, newspapers, periodicals) for readers at a 75 percent comprehension level—a level that is challenging, but not frustrating (Schnick and Knickelbine, 2000).

Within the *assessment* area, Lexile measures can be used to inform a variety of tasks related to test development and reporting. First, and foremost, the reading demands of the AIMS DPA and AIMS HS can be compared with the reading abilities of the students in Arizona and the reading demands of other similar assessments. Appendix A contains the readability analyses of the AIMS DPA and AIMS HS reading assessments.

Performance standards provide a common meaning of test scores throughout a state or nation concerning what is expected at various levels of competence. In Arizona, four performance standards have been established to describe reading performance at each grade level—Exceeds the Standard, Meets the Standard, Approaches the Standard, and Falls Far Below the Standard. Scoring at or above the cut score associated with “Meets the Standard” in grades 3 through 8 and High School is considered the standard representing satisfactory achievement of the knowledge and skills measured at the grade level. Just as Arizona has described what proficiency is, other states and publishers have also described this level. Below are the standards and descriptors for a norm-referenced assessment and a criterion-referenced assessment also linked to the Lexile scale.

For Nref1, standard setting was conducted using the task-centered, modified-Angoff method. Panelists were provided impact data concerning the percentage of students responding correctly to the item. Four achievement levels were established and the following descriptions of three of the standards were developed:

- Level 3, Proficient, represents *solid academic performance*, indicating that students are prepared for the next grade. At high school, this level reflects competency in a body of subject-matter knowledge and skills that prepares students for responsible adulthood and productive work.
- Level 2, Basic, denotes *partial mastery* of the knowledge and skills that are fundamental for satisfactory work. At the high school level, this is higher than minimum competency skills.
- Level 1, Below Basic, indicates *less than partial mastery*.

For State1, standard setting was conducted using the examinee-centered, contrasting-groups method. During the field test, teachers were asked to categorize each student participating in the field test into one of four proficiency levels. Teachers were asked to base their judgments on their firsthand knowledge of the student’s level of achievement during the school year in various domains assessed outside of the testing situation. Teachers are able to make informed judgments about students’ achievement because the teachers have observed the breadth and depth of the work each student has accomplished during the school year. Four achievement levels were established and the following descriptions of three of the standards were developed (emphasis added):

- Level III: Students performing at this level *consistently demonstrate mastery* of grade level subject matter and skills and are *well-prepared* for the next grade level.
- Level II: Students performing at this level *demonstrate inconsistent mastery* of knowledge and skills that are fundamental in this subject area and that are minimally sufficient to be successful at the next grade level.

- Level I: Students performing at this level *do not have sufficient mastery* of knowledge and skills in this subject area to be successful at the next grade level.

Initially it seems intuitive to compare the percentage of students that are above a particular standard in one state with similar results on another assessment. As can be seen from the standards described above there are many differences that affect the interpretations that can be made:

- Differences in *intent*—to describe what students “can do” as compared with what students “should be able to do;”
- Differences in *language*—“minimally proficient,” “does not have sufficient mastery,” or “partial mastery;” and
- Differences in *purpose*—to make high-stakes decisions about students or to describe performance on a domain.

Each of these differences impacts the valid comparisons that can be made between the results from two different assessments. The Lexile Map can be used to describe the reading demands on various assessments and provides an invariant frame of reference to make comparisons between assessments. The description of a AIMS DPA performance standard of “Meets the Standard” appears to correspond to the cut points between performance standards 2 and 3 on the assessments described above. This information can be used to compare the reading demands associated with the performance standards for the AIMS DPA with those associated with nationally-normed assessments and state-level criterion-referenced assessments.

Within the *communication* area, Lexile measures can be used to communicate with students, parents, teachers, educators, and the community by providing a common language to use to talk about reading growth and development.

By aligning all areas of the educational system, parents can be included in the instructional process. With a variety of data related to a student’s reading level a more complete picture can be formed and more informed decisions can be made concerning reading-group placement, amount of extra instruction needed, and promotion/retention decisions.

It is much easier to understand what a national percentile rank of 50<sup>th</sup> means when it is tied to the reading demands of book titles that are familiar to adults. Parents are encouraged to help their children achieve high standards by expecting their children to succeed at school, communicating with their children’s teachers and the school, and helping their children keep pace and do homework. Through the customized reading lists and electronic database of titles, parents can assist their children in the selection of reading materials that are at the appropriate level of challenge and monitor the reading process at home.

In one large school district, the end-of-year testing results are sent home to parents in a folder. The folder consists of a Lexile Map on one side and a letter from the superintendent on the other side. The school district considers this type of material as “refrigerator-friendly.” They encourage parents to put the Lexile Map on the refrigerator and use it to monitor and track the reading progress of their child throughout the school year.

The community-at-large (business leaders, citizens, politicians, and visitors) sees the educational system as a reflection of the community. Through the reporting of assessment results (after all, that is what the community is most interested in—results), people can

understand what the community values and see the return for its investment in the schools and its children.

One way to involve the community is to work with the public libraries and local bookstores when developing reading lists. The organizations should be contacted early enough so that they can be sure that the books will be available. Often books can be displayed with their Lexile measures for easy access.

Many school districts make presentations to civic groups to educate the community as to their reading initiatives and how the Lexile Framework is being utilized in the school. Conversely, many civic groups are looking for an activity to sponsor and it could be as simple as “donate-a-book” or “sponsor-a-reader” campaigns.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arizona Department of Education. (2003). *Arizona Language Arts Standards*. Retrieved August 1, 2007 from <http://www.ade.state.az.us/SBTL/sdi/langarts.asp>.
- Arizona Department of Education. (2005, September). *Arizona Reading Standard Performance Level Descriptors*. Retrieved August 13, 2007 from <http://www.ade.state.az.us/AIMS/descriptors>.
- Arizona Department of Education. (2007, January). *Testing Accommodations: Guidelines for 2006-2007*. Retrieved March 20, 2007 from <http://www.ade.state.az.us/standards/aims/Administering/>.
- Bormouth, J.R. (1966). Readability: New approach. *Reading Research Quarterly*, 7, 79-132.
- Carroll, J.B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.
- Carver, R.P. (1974). Measuring the primary effect of reading: Reading storage technique, understanding judgments and cloze. *Journal of Reading Behavior*, 6, 249-274.
- Chall, J.S. (1988). "The beginning years." In B.L. Zakaluk and S.J. Samuels (Eds.), *Readability: Its past, present, and future*. Newark, DE: International Reading Association.
- Crain, S. & Shankweiler, D. (1988). "Syntactic complexity and reading acquisition." In A. Davidson and G.M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Erlbaum Associates.
- CTB/McGraw-Hill. (2007, January). *Arizona: Arizona's Instrument to Measure Standards, 2006 Technical Report*. Author: Monterey, CA. Retrieved February 5, 2007 from <http://www.ade.state.az.us/standards/aims/Administering/Default.asp>.
- Davidson, A. & Kantor, R.N. (1982). On the failure of readability formulas to define readable text: A case study from adaptations. *Reading Research Quarterly*, 17, 187- 209.
- Dunn, L.M. & Dunn, L.M. (1981). *Manual for Forms L and M of the Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L.M. & Markwardt, F.C. (1970). *Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.
- Grolier, Inc. (1986). *The electronic encyclopedia*. Danbury, CT: Author.
- Klare, G.R. (1963). *The measurement of readability*. Ames, IA: Iowa State University Press.

- Kolen, M.J. & Brennan, R.L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.
- Kolen, M.J. & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and practices (Second Edition)*. New York: Springer Science+Business Media, Inc.
- Lieberman, I.Y., Mann, V.A., Shankweiler, D., & Westelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. *Cortex*, 18, 367-375.
- McGraw-Hill Book Company. (n.d.). *Guidelines for bias-free publishing*. Hightstown, NJ: Author.
- Miller, G.A. & Gildea, P.M. (1987). How children learn words. *Scientific American*, 257, 94-99.
- National Center for Education Statistics. (no date). NAEP -- State profiles: Arizona. Retrieved January 15, 2006, from <http://nces.ed.gov/nationsreportcard>.
- National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, D.C.: National Academy Press.
- Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). "Scaling, Norming, and Equating." In R.L. Linn (Ed.), *Educational Measurement* (Third Edition) (pp. 221-262). New York: American Council on Education and Macmillan Publishing Company.
- Poznanski, J.B. (1990). A meta-analytic approach to the estimation of item difficulties. Unpublished doctoral dissertation, Duke University, Durham, NC.
- Schnick, T. & Knickelbine, M. (2000). *The Lexile Framework: An introduction for educators*. Durham, NC: MetaMetrics, Inc.
- Shankweiler, D. & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 14, 139-168.
- Stenner, A.J. (1990). Objectivity: Specific and general. *Rasch Measurement Transactions*, 4, 111.
- Stenner, A.J., Smith, M., & Burdick, D.S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement*, 20(4), 305-315.
- Stenner, A.J., Smith, D.R., Horiban, I., & Smith, M. (1987a). Fit of the Lexile Theory to item difficulties on fourteen standardized reading comprehension tests. Durham, NC: MetaMetrics, Inc.
- Stenner, A.J., Smith, D.R., Horiban, I., & Smith, M. (1987b). Fit of the Lexile Theory to sequenced units from eleven basal series. Durham, NC: MetaMetrics, Inc.
- Tomlinson, C.A. (1999). *The differentiated classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wright, B.D. & Linacre, J.M. (1994, August). *The Rasch model as a foundation for the Lexile Framework*. Unpublished manuscript.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.

# Appendices

Appendix A: Readability of passages used with the AIMS DPA and AIMS HS ..... A-1

AZ Dual Purpose Assessment -- Lexile Measures

AZ Tests	Grade	Title	Words	MSL *	MLWF *	Lexile
<b>2007</b>	<b>Grade 3</b>	<b>All Passages</b>	3630	9.73	3.61722	620
<b>Each Passage</b>		1	49	9.80	3.37509	720
		2	341	11.76	3.71655	720
		3	359	11.58	3.94316	620
		4	168	9.88	3.23766	780
		5	479	8.12	3.94106	360
		6	492	10.47	3.77940	610
		7	195	12.19	3.52849	820
		8	277	6.76	3.74988	300
		9	131	8.19	3.54858	520
		10	382	11.94	3.69380	740
		11	327	8.84	3.79389	480
		12	293	9.77	3.64942	610
		13	137	10.54	3.53676	710
		14	NP	Poem		
	<b>Grade 5</b>	<b>All Passages</b>	5190	13.66	3.56332	890
<b>Each Passage</b>		1	567	14.18	3.70952	860
		2	625	12.76	3.76691	760
		3	235	13.82	3.53487	910
		4	245	12.25	3.41048	870
		5	518	12.63	3.64698	800
		6	357	15.52	3.49682	1010
		7	255	15.94	3.64844	970
		8	581	13.51	3.46659	920
		9	256	12.19	3.63103	780
		10	280	11.67	3.54872	780
		11	348	19.33	3.75522	1070
		12	561	14.03	3.61303	890
		13	362	12.93	3.46048	890
		14	NP			
		15	NP			
		16	NP			
		17	NP	Poem		
		18	NP	Poem		
		19	NP	Poem		

AZ Dual Purpose Assessment -- Lexile Measures

AZ Tests	Grade	Title	Words	MSL *	MLWF *	Lexile
	<b>Grade 7</b>	<b>All Passages</b>	7351	14.19	3.48198	950
<b>Each Passage</b>		1	236	14.75	3.57989	940
		2	489	12.87	3.45174	890
		3	390	11.14	3.51367	760
		4	401	11.79	3.21181	920
		5	293	17.24	3.13164	1230
		6	513	19.00	3.87575	1010
		7	249	14.65	3.13293	1110
		8	615	13.09	3.38095	930
		9	970	12.76	3.66602	800
		10	572	15.05	3.56691	960
		11	440	13.75	3.52488	910
		12	426	16.38	3.54717	1030
		13	573	15.49	3.59464	970
		14	597	16.14	3.62049	990
		15	113	18.83	3.50196	1150
		16	474	13.17	3.62227	840
		17	NP	Poem		
		18	NP			
		19	NP			
		20	NP			
		21	NP	Poem		
		22	NP			
		*	<b>MSL = Mean Sentence Length</b>			
			<b>MLWF = Mean Log Word Frequency</b>			

AZ High School -- Lexile Measures

AZ	Grade	Title	Words	MSL *	MLWF *	Lexile
	<b>Grade 10</b>	<b>All Passages</b>	6035	14.40	3.45876	970
		1	1144	16.58	3.67147	990
		2	283	17.69	3.60777	1180
		3	864	15.43	3.71521	920
		4	423	15.67	3.41216	1050
		5	414	18.00	3.26404	1210
		6	240	14.12	3.67641	870
		7	217	14.47	3.28834	1040
		8	728	11.03	3.54568	740
		9	320	14.55	3.32404	1030
		10	52	26.00	3.44278	1410
		11	NP	Poem		
		12	NP	Poem		
		13	NP	Poem		
		* MSL = Mean Sentence Length				
		MLWF = Mean Log Word Frequency				