

**Arizona English Language Learner Assessment
(AZELLA) Form AZ-2 Technical Report**

2009 – 2010 Administration

February 2011

Prepared by

**Pearson
San Antonio, TX**

Table of Contents

Summary of Technical Report.....	5
Classical Item-Level and Subtest Statistics.....	5
Reliability.....	5
Reliability of Classification Decisions at Proficient Cut.....	6
Validity.....	6
Calibration, Equating, and Scaling.....	7
IRT Statistics.....	7
Standard Setting.....	8
Summary of Operational Test Results.....	8
1. Introduction.....	9
1.1 Background.....	9
1.2 Rationale and Purpose.....	9
1.3 Recommended Test Use.....	10
1.4 Universal Design.....	10
1.5 Test Accommodations.....	10
Table 1.1: Test Accommodations for the Administration of the Arizona English Language Learner Assessment (AZELLA)	11
Source: Office of English Language Acquisition Services (April, 2010).....	11
2. Test Design and Development.....	12
2.1 Test Specifications by Subtest by Grade Span.....	12
Table 2.1: Test Specifications by Subtest by Grade Span	13
Table 2.2: Maximum Number of Points by Subtest by Grade Span	13
2.2 Item Development and Review Processes.....	13
2.3 Scoring of the Writing Forms Field Test.....	14
2.4 Test Construction.....	14
3. Data Sources and Item-Level Statistics.....	17
3.1 Data Sources.....	17
3.2 Item-Level Descriptive Statistics.....	17
3.3 Subtest Statistics.....	17
Table 3.1: Summary Statistics of Subtests by Grade Span (Forms Field Testing)	19
4. Reliability.....	20
4.1 Internal Consistency Reliability.....	20
4.2 Classical SEM (based on Classical Test Theory).....	20
4.3 Conditional SEM (based on Item Response Theory).....	21
4.4 Reliability of Each of the Reporting Strand.....	22
Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand	23
Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand (continued)	24
Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand (continued)	25
Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand (continued)	27
4.5 Reliability of Classification Decision at Proficient Cut.....	28
Figure 4.1: Classification Accuracy.....	28
Figure 4.2: Classification Consistency.....	28

Table 4.2: Decision Accuracy and Consistency by Grade	29
5. Validity	30
5.1 Test Content.....	30
5.2 Internal Structure	30
5.3 Evidence of the Internal Structure of AZELLA	30
Table 5.1: Intercorrelations Among the Subtests by Grade	32
5.4 Evidence of the Unidimensionality of the AZELLA	35
Table 5.2: Results of Principal Components Analysis	35
5.5 Relationships to Other Variables.....	36
Table 5.3: Statewide Comparisons of the Percent of Students Passing AIMS	37
6. Calibration, Equating, and Scaling	38
6.1 The Rasch and Partial Credit Models	38
Figure 6.1: Sample Item Characteristic Curve	39
Figure 6.2: Category Response Curves for a Dichotomous or One-step Item	40
Figure 6.3: Category Response Curves for a Two-step Item	41
6.2 Calibration, Equating, and Scaling of the AZELLA Form AZ-2.....	42
Table 6.1: Number of Linking Items	43
6.4 Scaled Scores.....	43
7. IRT Statistics	44
7.1 Model and Rationale for Use.....	44
7.2 Evidence of Model Fit.....	44
7.3 Rasch-Based Item Information.....	45
Table 7.1: Average Rasch Difficulty by Grade Span by Subtest	45
Table 7.1: Average Rasch Difficulty by Grade Span by Subtest (Continued)	46
8. Standard Setting.....	47
8.1 Introduction	47
8.2 Proficiency Categories for AZELLA	47
8.3 Composition of Standard Setting Committees	47
Table 8.1: Panel Composition for Standard Setting Committees	48
8.4 The Standard Setting Process	50
8.5 Introduction to the Process	50
8.6 Independent Ratings of Each Item	51
8.7 Provision and Discussion of Data.....	51
8.8 Adjustment of Judges' Cut Scores	52
8.9 Analyses After Standard Setting	52
Table 8.8a: Scaled Score Cut Ranges (Grades K–5)	53
Table 8.8b: Scaled Score Cut Ranges (Grades 6–12)	55
9. AZELLA Form AZ-2 2009-2010 Administration Results.....	57
9.1 Results Based on All Student Records	57
9.2 Results Without Reclassified English Language Learners Students	57
Table 9.1: Percentage of Students in Each Proficiency Category	59
Table 9.2: Raw Score Descriptive Statistics by Grade	62
Table 9.3: Scaled Score Descriptive Statistics by Grade	65
Table 9.4: Descriptive Statistics for the Total Score without Reclassified Students by Grade	68
Table 9.5: Descriptive Statistics for the Total Score without Reclassified Students by Grade and Time of Year	69

REFERENCES..... 70

SUMMARY OF TECHNICAL REPORT

This is a summary of the technical information for the Arizona English Language Learner Assessment (AZELLA) Form AZ-2. It summarizes essential psychometric information including:

- Classical item-level and subtests statistics;
- Reliability;
- Validity;
- Calibration, equating, and scaling;
- IRT statistics;
- Standard setting; and
- Summary of the operational test results for the 2009-2010 administration.

Classical Item-Level and Subtest Statistics

This section presents the item-level and subtest summary statistics using raw scores under classical test theory.

Appendix A shows that in general, the item difficulties are moderately easy and none of the items had a negative point biserial across all levels of the test.

The subtest statistics are also presented by the reporting strands. The classical measures of central tendency, variability, and score precision are reported for each of the reporting strands. The mean, standard deviation, and standard error of mean are presented in Table 3.1 (page 19) (based on the Forms Field Test Fall 2008).

Reliability

Reliability is the degree to which scores remain consistent over an assessment procedure (Nitko, 2004). Further defined, reliability is the degree to which students' assessment results are consistent when a) students complete the same task on two or more occasions, b) two or more raters evaluate their performance on the same task, or c) students complete two or more parallel tasks on one or more occasions. Consistency of scores over repeated assessment and/or with different raters is the underlying feature of reliability.

To report and document the reliability of the AZELLA Form AZ-2, three indices were used:

- The internal consistency—Cronbach's coefficient alpha (Cronbach, 1951);
- The standard error of measurement (SEM) (based on Classical Test Theory);
- The conditional SEM (based on Item Response Theory).

Table 4.1 (page 21) provides reliabilities by grade and by strand. The reliability of the Total Composite for grade K is 0.80 and for grades 1 through 12 are in the high 80s and low 90s.

The conditional SEMs are presented in the raw score-to-scaled score conversion tables in Appendix B. The patterns of conditional SEMs across grade spans are reasonable, as the smaller values occur at the middle of the scale.

Reliability of Classification Decisions at Proficient Cut

Based on the AZELLA Form AZ-2 scaled scores, student performance is classified into one of five proficiency levels. While it is always important to know the reliability of student scores in any examination, it is of even greater importance to assess the reliability of the decisions based on these scores. Evaluation of the reliability of classification decisions is performed through estimation of the probabilities of correct and consistent classifications of student performance. Procedures from Livingston and Lewis (1995) were applied to derive measures of the accuracy and consistency of the classifications.

Accuracy: The accuracy of decisions is the extent to which decisions would agree with those that would be made if each student could somehow be tested with all possible forms of the examination.

Consistency: The consistency of decisions is the extent to which decisions would agree with the decisions that would have been made if the students had taken a parallel form of the AZELLA Form AZ-2, equal in difficulty and covering the same content as the form they actually took.

Estimates of decision accuracy and consistency were made for the “Achieves Proficient Status” cut point on the Total Composite score reported in the AZELLA Form AZ-2, shown in Table 4.2 (page 29).

Table 4.2 also includes the proportions of False Positive and False Negative classifications. False Positive and False Negative classifications refer to the mismatch between students’ true scores and observed scores.

False Positive: The False Positive value is the proportion of student scores misclassified to the category “Achieves Proficient Status” when student scores do not meet proficient status.

False Negative: The False Negative value is the proportion of student scores misclassified to the category “Does Not Achieve Proficient Status” when student scores actually do meet proficient status.

The sum of values of Accuracy, False Positive, and False Negative is equal to 1.00.

Table 4.2 presents the results of the decision accuracy and consistency of the “Achieves Proficient Status” cut score for the Total Composite scores. The result illustrates the general rule that decision consistency will be less than the decision accuracy. It should also be noted that the decision accuracy of students for students who achieved Proficient Status for the Total Test ranged from 0.89 to 0.96.

Validity

For the AZELLA Form AZ-2, Form C of the Stanford English Language Proficiency Test (SELP) was used together with augmented items created by Arizona teachers to construct one form per grade span. Special calibration studies were conducted in order to obtain both traditional and Rasch item statistics. Section 6 of this manual describes the calibration, equating, and scaling procedures conducted on the Forms Field Test Fall 2008 dataset. A wealth of item information was gathered through these calibration studies. Among the statistics included are *p*-values, point-biserials, Rasch difficulty, and standard error of the Rasch difficulty. The AZELLA Form AZ-2 supports the validity-related standards set forth in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Judgments about test validity are based on the following sources of evidence of validity:

- Test content—“an analysis of the relationship between a test’s content and the construct it is intended to measure” (AERA, APA, NCME, 1999, p. 11.)

- Internal structure—“the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are made” (AERA, APA, NCME, 1999, p. 13)

Table 5.1 presents the intercorrelations among the four subtests by grade. The results of this analysis showed that the correlations between the Reading and Writing subtests were consistently the highest compared to the other combinations of subtests for each grade.

The evidence of internal structure is also depicted by the point biserial correlation coefficient and fit statistics. Appendix A provides these statistics.

- Relationships to other variables—“analyses of the relationship of test scores to variables external to the test” (AERA, APA, NCME, 1999, p. 13)

Evidence of validity based on relationships to other variables will be conducted as ongoing research. For example, data from the 2010 administration of the Arizona’s Instrument to Measure Standards (AIMS) should be obtained and matched to AZELLA AZ-2 data in order to examine the relationship between English language proficiency and student achievement.

Calibration, Equating, and Scaling

Pearson used the pre-existing SELP vertical scale to create the AZELLA vertical scale. SELP items, which comprised about 30% of the items on the AZELLA Form AZ-2, were fixed to the parameter values from the pre-existing vertical scale. Any remaining non-SELP items on the AZELLA Form AZ-2 were calibrated together with the SELP items using the Rasch and Partial Credit model. By fixing the values of the SELP items prior to calibration, the item difficulty and step parameters of all the items were placed on the same vertical scale.

Table 6.1 (page 39) shows the number of original linking items and the total number of final linking items used.

Appendix B provides the raw score-to-scaled score conversion tables for the reporting strands by grade span.

IRT Statistics

Fit statistics are used for evaluating the goodness-of-fit of a model to the data and are calculated by comparing the observed and expected trace lines obtained for an item. Two forms of fit statistics, OUTFIT (relatively sensitive to outliers, or highly unexpected responses) and INFIT (relatively sensitive to patterns of misfit), are reported. The cutoffs are set at less than 0.5 and greater than 1.5.

The OUTFIT and the INFIT statistics are presented in the item statistics tables in Appendix A. Using the cutoffs described above, there were 4% of the Middle Grades items flagged for INFIT. The percentage of items that were flagged for OUTFIT ranged from 1% to 5%.

Table 7.1 (page 45) reports the average Rasch difficulty by grade span by subtest. Appendix A presents the Rasch information at the item level.

Standard Setting

Standard setting was held for AZELLA Form AZ-1. There was not a separate standard setting for the AZELLA Form AZ-2. The AZELLA Form AZ-2 cut points were equated to the original AZELLA cut points. Following is the description of the standard setting meeting conducted on the AZELLA Form AZ-1.

As the contractor for the AZELLA, Harcourt Assessment¹ organized a performance standard setting meeting. Harcourt Assessment involved 77 Arizona educators in a two-day standard setting meeting. The purpose of the meeting was to obtain preliminary recommendations for the AZELLA cut scores for five performance levels (Pre-Emergent, Emergent, Basic, Intermediate, and Proficient) for each of five grade bands (K, 1–2, 3–5, 6–8, 9–12). A modified-Angoff procedure was used. After completion of the standard setting meeting, Harcourt Assessment conducted several post-standard setting analyses and developed the approved AZELLA cut score ranges in scaled score for the reporting strands for all grades. Tables 8.8a (page 53) and 8.8b (page 55) show the cut score ranges.

Summary of Operational Test Results

The raw score and scaled score summaries are reported in the following tables for each of the reporting strands. These data are taken from the 2009-2010 AZELLA AZ-2 operational administration and are based on all student data (pages 59-69):

Table 9.1 presents the percentages of students in proficiency category by grade,

Table 9.2 presents the raw score descriptive statistics by grade, and

Table 9.3 presents the scaled score descriptive statistics by grade.

Additional tables are presented based on student data where English language learners deemed proficient through assessment with AZELLA were reclassified as “fluent English proficient” and removed from the data:

Table 9.4 shows this for the Total Test scores aggregated across the 2009-2010 school year.

Table 9.5 shows the results for the Total Test scores broken out by time of year (i.e., Fall, Mid-year, or Spring).

¹ Harcourt Assessment was acquired by Pearson in January 2008.

1. INTRODUCTION

1.1 Background

Title III of the federal *No Child Left Behind* (NCLB) Act of 2001 requires annual assessment of the English proficiency of limited English proficient students. NCLB requires demonstrated annual improvement and adequate yearly progress for such students in order for them to develop English proficiency and meet challenging State academic content standards. Arizona state law also requires annual reassessment of limited English proficient students using a state-approved assessment.

To meet these requirements, the Arizona Department of Education (ADE) requested the research, development, and scoring of the five grade spans and four subtests linked to the State's approved K–12 English Language Learner (ELL) proficiency standards. The test was developed for five grade spans (K, 1–2, 3–5, 6–8, and 9–12) and in four subtests (Listening, Speaking, Reading, and Writing) to assess the English language proficiency of students in kindergarten through grade 12 who are English language learners. The *Arizona English Language Learner Assessment* (AZELLA Form AZ-1) was developed in conjunction with Aha Inc., a state contractor for facilitation, consulting, and technical writing services. The assessment was developed in accordance to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and Arizona State testing requirements. It is consistent with the principles of universal design and applicable federal and state testing requirements.

AZELLA Form AZ-1 became the operational assessment for use with ELL students in fall 2006 and has continued in use through spring 2009.

For the 2009-2010 test administration, Pearson proposed the use of the *Stanford English Language Proficiency Test* (SELP), Form C, in conjunction with new items written by a team of Arizona educators. This second version of the *Arizona English Language Learner Assessment* (AZELLA) is AZELLA Form AZ-2.

1.2 Rationale and Purpose

The ADE has established proficiency standards for all English language learners attending Arizona schools. In compliance with NCLB, the Department developed a test that measures student progress toward meeting these standards. This test is the *Arizona English Language Learner Assessment* (AZELLA Form AZ-2). NCLB mandates that all English language learners from kindergarten through grade 12 be assessed every year to measure their English language proficiency in listening, speaking, reading, and writing, and to track their annual progress toward proficiency. In addition, Arizona state law requires a reassessment of English language learners at least annually at the end of each school year. AZELLA assists educators by identifying students' language performance capabilities on the proficiency standards. Test results provide specific information that can be used to target instruction and ensure that English language learners fully acquire the language proficiency needed for educational success.

The purposes of AZELLA are (1) to determine appropriate placement of students who have a Primary Home Language Other Than English (PHLOTE) and (2) to measure ELL students' annual improvement in achieving English language proficiency. Thus, the test results provide the criteria for entry into ELL

programs as well as for exiting ELL programs. English language learners deemed proficient through assessment with AZELLA are reclassified as “fluent English proficient” and exit the ELL program. These students are then placed into mainstream classes. Reclassified English language learners are tested annually at the end of each school year for two years. Students who fail to demonstrate proficiency are eligible to be reenrolled in ELL programs.

1.3 Recommended Test Use

AZELLA is designed to assess students at all proficiency levels within each grade span. This vertical development of the language tested allows the test to discriminate more finely among students at different stages of language acquisition. Because test results provide students, teachers, and parents with an objective report of each student’s strengths and weaknesses in the English language skills of listening, speaking, reading, and writing, AZELLA will help determine whether students are making adequate progress toward English language proficiency. Year-to-year progress in language proficiency can also be measured and documented. The test results also will help schools focus on ways to improve instruction so that English language learners become proficient in English.

1.4 Universal Design

All items were developed following the guidelines of universal design. Adherence to these guidelines ensured that the assessments were accessible and valid for the widest range of students, including students with disabilities. Every item was carefully reviewed to ensure it was built taking into consideration equitable use, flexibility in use, simple and intuitive design, perceptible information, tolerance for error, low physical effort, and size and span for approach and use. As forms were constructed, Pearson’s content and fairness experts reviewed the forms to ensure that concepts of universal design were incorporated. A review of forms for use by special populations, such as the visually or hearing-impaired student, was conducted to ensure that items were fair, reliable, and accessible to all.

1.5 Test Accommodations

Table 1.1 shows the test accommodations that are allowed for the administration of the AZELLA. These assessment accommodations include those for timing/scheduling, setting/administration, presentation format, and response format. While most accommodations are considered acceptable for the administration of AZELLA, there are some accommodations that are not acceptable such as the use of translations, dictionaries, or glossaries.

Table 1.1: Test Accommodations for the Administration of the Arizona English Language Learner Assessment (AZELLA)

Accommodations	Acceptable	Not Acceptable
Timing/Scheduling		
• Breaks between subtests	X	
• Time of day most beneficial to student	X	
• Frequent breaks within a subtest	X	
• Time as needed	X	
Setting/Administration		
• Test in a small group	X	
• Test individually	X	
• Environmental modification: Location with minimal distractions, preferential seating, noise buffers	X	
Presentation Format		
• Repeating directions in English	X	
• Simplifying directions in English	X	
• Color overlay for use by visually impaired students	X	
• Use of hearing devices for hearing impaired students: hearing aids or amplification	X	
• Reading aloud any test questions or prompts when not required by test directions		X
Response Format		
• Use of special pencil or pencil grip	X	
• Scribe to assist in cases of injury such as broken arm or hand for bubble answers (e.g., multiple choice)	X	
• Scribe to assist in cases of injury such as broken arm or hand for written responses	X	
• Oral response in native language		X
Other		
• Translation into any language		X
• Use of dictionaries		X
• Use of word list/glossaries		X
Test Modifications		
• Enlargement of test material for use by visually impaired students	X	

Source: Office of English Language Acquisition Services (April, 2010).

2. TEST DESIGN AND DEVELOPMENT

2.1 Test Specifications by Subtest by Grade Span

AZELLA has four subtests (Speaking, Listening, Reading, and Writing) for grades K–12. It includes multiple-choice, constructed-response, short-response, and extended-response items. As shown in the chart on the next page, the numbers of items per grade span increases from K to 12. Grade K has the least number of items with 53, grades 1–2 and 3–5 have 76 items, and grades 6–8 and 9–12 have 84 items.

The Speaking subtest has 12 constructed-response items for the grade K grade span and 14 constructed-response items for all other grade spans. The Listening subtest has only multiple-choice items. The number of items for the Listening subtest increases from 12 on the grade K test to 20 for the rest of the grade spans. The number of items for the Writing subtest ranges from 17–26 for the various grade spans. The Writing subtest has three parts:

- Multiple-choice section that assesses English language learners' understanding of standard English conventions
- Pre-writing items (kindergarten only)
- Two extended responses to graphics-based prompts for grades 1 through 5; one extended response to graphics-based prompts and one extended response without graphic-based prompts for grades 6 through 12

For kindergarten, which is usually individually administered, the Listening and Pre-reading subtests are multiple-choice. Pre-writing consists of 17 early writing production items and Speaking has constructed oral response items.

The test design for the AZELLA Form AZ-2 is shown in Table 2.1. Table 2.2 provides the maximum number of points by subtest by grade span. This design has items from the SELP, Form C, and new items written by a team of Arizona educators.

Table 2.1: Test Specifications by Subtest by Grade Span

Number of Items and Passages in AZELLA Form AZ-2 Subtests								
Grade Span	Speaking	Listening	Reading		Writing			Total Number of Items per Grade Span
	CR	MC	MC	Passages	Writing Conventions MC	Pre-Writing SR	Writing Prompt ER	
K	12	12	12	0	0	16	1	53
1–2	14	20	20	10	20	0	2	76
3–5	14	20	20	5	20	0	2	76
6–8	14	20	24	4	24	0	2	84
9–12	14	20	24	4	24	0	2	84

Table 2.2: Maximum Number of Points by Subtest by Grade Span

Maximum Number of Points in AZELLA Form AZ-2 Subtests								
Grade Span	Speaking	Listening	Reading		Writing			Total Number of Points per Grade Span
	CR	MC	MC	Passages	Writing Conventions MC	Pre-Writing SR	Writing Prompt ER	
K	26	12	12	0	0	20	2(SR)	72
1–2	32	20	20	10	20	0	8	100
3–5	32	20	20	5	20	0	8	100
6–8	32	20	24	4	24	0	8	108
9–12	32	20	24	4	24	0	8	108

2.2 Item Development and Review Processes

In order to create a new and fully aligned assessment for English language learners for the 2009-2010² (Form AZ-2) administration, and also to meet the reporting requirements for NCLB, Pearson made use of the SELP, Form C, and worked in conjunction with a team of Arizona educators to produce customized items. The Arizona Department of Education (OELAS division) reviewed the AZELLA AZ-2 forms prior to administration.

² AZELLA is administered throughout the academic year.

2.3 Scoring of the Writing Forms Field Test

Writing field test student responses were reviewed, and a low, medium, and high range set of ten papers each and two mixed sets of ten papers each were identified in advance of range finding for each of the two new operational prompts in primary and elementary levels and for the one new prompt in middle and high school levels. From these student responses, three examples of appropriate anchor papers that represented each of the score points of 0 through 4 were selected during range finding and anchor pulling activities, participated in by Arizona educators and facilitated by Pearson Performance Scoring Center (PSC) personnel, held in Arizona. After the anchors were selected, the remaining responses were used for training sets and qualifying sets. Legacy anchors for the SELP, Form C prompts in middle and high school were reviewed and approved with a few substitutions made from the forms field test responses.

Prior to scoring live student documents, raters reviewed the anchors and the accompanying annotations. Thereafter, each rater examined the papers in Training Set A and assigned scores. The scores were then verified and the rater had an opportunity to discuss the scores with the Project Scoring Director. This process was repeated for Training Set B. The rater was then ready to qualify to score. Raters were required to take two qualifying sets containing 10 papers each, and pass at least one qualifying set with a score of 80% correct. If they did not do this, they were not allowed to score any live student responses

During the scoring of live student responses, the Scoring Director second scores 20% of each rater's work product. The agreement rate between the two raters' scores is tracked by the Scoring Director. Feedback about accuracy and productivity is given to the raters every morning by the Scoring Director. Additionally, raters can ask scoring questions of the Scoring Director.

2.4 Test Construction

At least 30 % of SELP, Form C items were retained in each domain (Reading, Writing and Writing Conventions, Listening, and Speaking). Arizona educators developed new test items that were grade-band appropriate. They used the reformatted Arizona K–12 English Language Learner Proficiency standards to ensure that the new test items measured the skills required in specific standards' proficiency indicators. These items were reviewed by a Bias and Content Review Committee in September 2008. Several items were modified, deleted, or replaced before all the test items were submitted to Pearson for their review and editing. The revision process continued for several months. The finalized test questions were item field-tested. After the results of the item field tests were analyzed, the final operational forms field test was developed by Pearson.

The AZELLA Form AZ-2 represents a broad range of difficulty at all grade spans from K–12. A broad range of test items includes some very simple test items that have high *p*-values for students with little or no ability in English as well as test items that have low *p*-values for students with advanced ability in English.

Testing Written Language

A fundamental consideration in constructing AZELLA is the language that is being assessed. While this question can generally be answered from the test developer's native speaker intuition, more rigorous

methods in language choice need to be applied to provide consistency across the forms of the five grade spans and to create a vertical structure within each form. Vertical structure is language that ranges from most simple language first acquired by non-native speakers to advanced language that indicates a level of English proficiency sufficient for participation in regular academic classes.

The vertical structure of AZELLA allows the assessment to discriminate more finely among students at different stages of language acquisition. The accurate identification of students at different levels of language development provides critical information to classroom teachers who can develop and apply effective instructional strategies to help their students reach proficiency. In addition to determining language proficiency, AZELLA provides evidence of students' progress toward proficiency required by the *No Child Left Behind Act* of 2001.

To determine the appropriate language for English language learner items and stimuli, Pearson assessment specialists, editors, and item and passage writers apply the Flesch-Kincaid grade span readability measures to all reading passages. Readability measures are primarily based on factors such as the number of words in the sentences and the number of letters or syllables per word. ESL assessment specialists also evaluate the coherence, the number of anaphora, vocabulary difficulty, sentence and text structure, and concreteness and abstractness of a passage. The sum of these elements determines the appropriateness of the language of a passage.

All grade spans of AZELLA contain multiple reading passages, and these passages increase in difficulty both within the grade span test and among the grade span tests. Pearson also uses the *Educational Developmental Laboratory (EDL) Core Vocabularies in Reading, Mathematics, Science, and Social Studies*, published by Steck-Vaughn, to help determine age- and grade-appropriate language for English language learner items and stimuli for the oral language subtests. Pearson ESL assessment specialists and editors ensure that the language in all stimuli and items, from kindergarten through grade 12, is both topic- and age-appropriate for test takers.

Testing Oral Language

Recognizing that oral language structure and vocabulary of English differ vastly from the written language, issues of oral language assessment among kindergarten through grade 12 English language learners were studied in 2005. In that year, Harcourt Assessment's English language proficiency professionals individually administered a pilot of the Listening and Speaking tests to English language learners, observing and recording student responses and eliciting their reactions to the tests. Outcomes of the pilot led to design decisions regarding:

- Item types;
- Numbers of items;
- Length of pauses between items;
- Use of recorded stimuli;
- Recording student spoken responses.

The Listening and Speaking subtests of AZELLA are based on these decisions. To ensure that the language in the Listening and Speaking stimuli and items reflect current spoken language as much as possible,

Listening and Speaking scripts are submitted to a read-aloud proofing process with English language learner assessment specialists and editors. Additionally, for the oral components of AZELLA to be relevant, the Listening and Speaking tests must have predictive validity for academic achievement; therefore, academic language as well as social language is an integral part of the Listening and Speaking subtests of AZELLA.

3. DATA SOURCES AND ITEM-LEVEL STATISTICS

3.1 Data Sources

There are three main sources of data used in the analyses of this technical report. The first is based on the AZELLA Form AZ-2 Forms Field Test data. The second source of data is from the 2009-2010 operational administration excluding students with missing subtest scores. The third source of data is from the full 2009-2010 operational administration of the AZELLA including students who may have taken the test more than once during the school-year.

3.2 Item-Level Descriptive Statistics

This section presents the raw score summary statistics for all items in the AZELLA Form AZ-2 that were tested between October 14 and November 21, 2008 (Forms Field Test). The p -value for each item is defined as the proportion of students that answer an item correctly for the multiple-choice items. A high p -value means that an item is easy; a low p -value means that an item is difficult. For the constructed-response items, the p -value is reported as the average number of points out of the maximum number of possible points.

The point biserial correlation for each item is an index of the association between the item-score and the total-test score. It shows the ability of the item to discriminate between low- and high-scoring students. An item with a high point biserial correlation discriminates more effectively between the low- and the high-scoring students than a low point biserial correlation.

The item-level statistics for the Forms Field Testing are presented in Appendices A.1–A.5 by grade span. The tables are grouped by the following subtests: Listening, Writing Conventions, Reading, Writing Open-Ended, and Speaking. The following item information and statistics are presented for each item:

- Item number;
- Item format (multiple-choice; constructed-response; short-response; or extended-response);
- Maximum number of possible points;
- N-count (number of students);
- p -value for multiple-choice items (percentage of examinees that answered the item correctly);
- Item mean for constructed-response items (average number of points earned out of the maximum number of possible points);
- Point biserial (index of discrimination between high- and low-scoring students)
- Rasch item difficulty;
- Standard error of Rasch difficulty;
- INFIT and OUTFIT statistics.

3.3 Subtest Statistics

The AZELLA Form AZ-2 scores based on the Forms Field Testing are reported for the following strands: Listening, Writing Conventions, Reading, and Speaking. There are four additional subtest scores reported for AZELLA based on combinations of items from these: Comprehension is based on items from Listening and Reading; Oral is based on items from Listening and Speaking; Total Writing is based on items from Writing Conventions and Writing Open-Ended; and Total Test is based on items from Listening, Reading, Speaking

and Total Writing. The classical measures of central tendency, variability, and score precision are reported for these reporting strands. The mean, standard deviation, and standard error of mean (SE of Mean) are presented in Table 3.1 by the subtests. The table includes the following:

- Level (grade span);
- Strand;
- Maximum score attainable;
- Maximum point received;
- N-count (sample size);
- Mean (average raw score);
- SD (standard deviation of raw scores); and
- SE of mean (standard error of mean).

Table 3.1: Summary Statistics of Subtests by Grade Span (Forms Field Testing)

Level	Strand	Max Points Possible	Max Points Received	N-Count	Mean	SD	SE of Mean
Preliteracy K	Listening	12	12	735	9.34	1.72	0.06
	Reading	12	12	735	8.08	2.46	0.09
	Speaking	26	26	735	18.03	5.83	0.22
	Comprehension	24	24	735	17.42	3.56	0.13
	Oral	38	38	735	27.37	6.88	0.25
	Total Writing	22	21	735	14.21	4.19	0.15
	Total Test	72	71	735	49.66	10.91	0.40
Primary 1–2	Listening	20	20	639	18.07	2.18	0.09
	Writing	20	20	639	13.62	4.58	0.18
	Reading	20	20	639	13.08	4.24	0.17
	Speaking	32	32	639	24.42	5.89	0.23
	Comprehension	40	40	639	31.16	5.52	0.22
	Oral	52	52	639	42.50	7.01	0.28
	Total Writing	28	28	639	17.81	5.81	0.23
Total Test	100	100	639	73.39	14.44	0.57	
Elementary 3–5	Listening	20	20	1004	14.58	2.84	0.09
	Writing	20	20	1004	14.63	3.71	0.12
	Reading	20	20	1004	10.89	3.77	0.12
	Speaking	32	32	1004	27.51	4.65	0.15
	Comprehension	40	38	1004	25.47	5.79	0.18
	Oral	52	51	1004	42.09	6.25	0.20
	Total Writing	28	27	1004	19.41	4.51	0.14
Total Test	100	95	1004	72.39	11.96	0.38	
Middle Grades 6–8	Listening	20	20	1076	15.17	3.23	0.10
	Writing	24	24	1076	18.78	4.15	0.13
	Reading	24	24	1076	16.98	4.63	0.14
	Speaking	32	32	1076	26.23	6.04	0.18
	Comprehension	44	44	1076	32.14	7.01	0.21
	Oral	52	52	1076	41.40	8.03	0.24
	Total Writing	32	32	1076	23.76	4.79	0.15
Total Test	108	106	1076	82.13	14.76	0.45	
High School 9–12	Listening	20	20	1489	15.22	3.48	0.09
	Writing	24	24	1489	17.53	3.93	0.10
	Reading	24	24	1489	16.21	4.38	0.11
	Speaking	32	32	1489	24.02	7.53	0.20
	Comprehension	44	44	1489	31.43	7.08	0.18
	Oral	52	52	1489	39.24	10.08	0.26
	Total Writing	32	32	1489	23.11	4.83	0.13
Total Test	108	106	1489	78.56	16.99	0.44	

4. RELIABILITY

Reliability is the degree to which scores remain consistent over an assessment procedure (Nitko, 2004). Further defined, reliability is the degree to which students' assessment results are consistent when a) students complete the same task on two or more occasions, b) two or more Raters evaluate their performance on the same task, or c) students complete two or more parallel tasks on one or more occasions. Consistency of scores over repeated assessment and/or with different Raters is the underlying feature of reliability.

4.1 Internal Consistency Reliability

The internal consistency of a test investigates the stability of scores from one sample of content to another. Several methods can be used to estimate the internal consistency of a test. One approach is to split all test questions into two groups and then correlate student scores on the two half-tests. This is known as a split-half estimate of reliability. This method avoids the implications of any changes in the individual by administering only a single test. If scores have a high rate of correlation on the two half-tests, it can be concluded that the test questions complement one another, function well as a group, and measure similar concepts. This also suggests that measurement error is minimal.

The split-half method's decision about which questions contribute to each half-test's score can have an impact on the resulting correlation. Pearson uses Cronbach's coefficient alpha statistic (Cronbach, 1951) to avoid this concern about the split-half method. The coefficient alpha is the average split-half correlation based on all possible divisions of a test into two parts. The coefficient alpha can be used to estimate the internal consistency of both dichotomously (right or wrong, 0 or 1 score values) and polytomously (a wide range of score values) scored test items. Coefficient alpha is computed by the following formula:

$$\alpha = \frac{I}{I-1} \left(1 - \frac{\sum_i s_i^2}{S_x^2} \right)$$

where

I is the number of items on the test,

s_i^2 is the variance of item i , and

S_x^2 is the total test variance.

4.2 Classical SEM (based on Classical Test Theory)

Since no assessment measures ability with perfect consistency, it is useful to take into account the likely size of measurement errors. One way to describe the inconsistency of assessment results is to assess a student on multiple occasions and note how much the scores vary. Repeatedly measuring a student can be done only hypothetically; however, if a student could be assessed on multiple occasions, a collection of the student's scores could be obtained. The scores would cluster around an average value. The standard deviation, or spread, of these obtained scores is known as the standard error of measurement (SEM).

The SEM is another index of reliability and provides an estimate of the amount of error in an individual's observed test score. The measurement error is commonly expressed in terms of standard deviation units; that is, the standard error of measurement is the standard deviation of the measurement error distribution. The standard error of measurement is calculated with the following equation:

$$SEM = SD\sqrt{1-r_{xx}}$$

where

SEM refers to the standard error of measurement,

SD is the standard deviation unit of the scale for a test, and

r_{xx} is the reliability coefficient for a sample test (or estimate of ρ_{XX} , which is a population reliability coefficient).

Note that the greater the reliability, the lower the standard error of measurement, and the more confidence one may have in the accuracy, or precision, of the observed test score.

4.3 Conditional SEM (based on Item Response Theory)

Unlike the SEM based on the classical test theory, the SEM based on the Item Response Theory (IRT) is not the same for all persons. For example, if a person gets either a few or a large number of items correct (extreme score), the standard error is greater than if the person gets a moderate number of items correct. This implies that the standard error of measurement depends on the total score (Andrich & Luo, 2004).

Under the Rasch model, the SEM for each person is as follows:

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^L p_{vi}(1-p_{vi})}}$$

where

v is subscript for a person,

i is subscript for an item,

L is length of the test,

$\hat{\beta}$ is the ability estimate, and

p_{vi} is the probability that a person answers an item correctly and is defined as follows:

$$P_{vi} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

where β_v is person v 's ability and δ_i is the item's difficulty.

Note that the standard error for item difficulty is smallest when the probability of passing is close to the probability of failing. That is, when an item is near the threshold level for many persons in the sample, the standard error is small (Embretson & Reise, 2000).

The conditional standard errors of measurement are presented in the raw score-to-scaled score conversion tables in Appendix B.

4.4 Reliability of Each of the Reporting Strand

Table 4.1 provides the raw score and scaled score descriptive statistics and reliabilities by grade by the reporting strands using the operational data from the 2009-2010 administration of AZELLA Form AZ-2 . Only students who took all subtests are included in the analyses. The following information is captured:

- Number of items;
- Maximum number of possible points;
- Number of students;
- Means and standard deviations in raw scores
- Means and standard deviations in scaled scores;
- Kuder-Richardson Formula 20 (KR20) internal consistency reliability estimate for the multiple-choice items and Cronbach's alpha for the extended-response items;
- Standard error of measurement;
- Mean raw score as a proportion of the maximum obtainable score; and
- IRT based conditional standard errors of measurement for the Proficient cut scores.

Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand

Kindergarten											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	12	12	40468	9.39	1.97	522.08	50.60	0.62	1.22	0.78	30.14
Prereading	12	12	40468	8.15	2.90	526.56	54.25	0.76	1.42	0.68	36.82
Speaking	12	26	40468	18.33	6.56	530.97	60.76	0.91	1.96	0.71	31.29
Comprehension	24	24	40468	17.55	4.39	523.46	48.25	0.81	1.89	0.73	25.62
Oral	24	38	40468	27.73	7.89	526.07	51.36	0.90	2.55	0.73	26.50
Prewriting	17	22	40468	13.99	6.16	527.74	88.16	-	-	0.64	21.18
Total Composite	36	72	40468	49.87	14.78	524.94	52.52	0.80	6.68	0.69	18.27
1st Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	23896	18.21	2.16	622.11	46.98	0.71	1.17	0.91	20.97
Writing	20	20	23896	14.43	4.08	627.08	50.37	0.82	1.74	0.72	22.33
Reading	20	20	23896	14.03	3.84	621.25	47.00	0.79	1.77	0.70	22.51
Speaking	14	32	23896	25.75	5.72	630.37	55.66	0.88	1.99	0.80	16.52
Comprehension	40	40	23896	32.24	5.18	620.14	44.46	0.82	2.17	0.81	16.66
Oral	34	52	23896	43.96	6.89	625.85	49.48	0.87	2.52	0.85	13.55
Total Writing	22	28	23896	19.06	5.25	627.73	50.31	0.76	2.55	0.68	20.62
Total Composite	76	100	23896	77.05	13.26	620.78	38.52	0.90	4.15	0.77	10.89
2nd Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	23788	19.10	1.44	644.94	39.46	0.63	0.88	0.95	27.06
Writing	20	20	23788	17.20	3.18	666.67	50.38	0.82	1.35	0.86	22.33
Reading	20	20	23788	17.02	3.10	664.32	50.81	0.80	1.40	0.85	26.57
Speaking	14	32	23788	27.86	4.79	654.96	57.20	0.87	1.75	0.87	17.78
Comprehension	40	40	23788	36.11	3.99	661.33	49.34	0.82	1.68	0.90	19.39
Oral	34	52	23788	46.96	5.53	652.82	52.93	0.85	2.13	0.90	14.98
Total Writing	22	28	23788	23.20	4.35	677.54	58.57	0.82	1.87	0.83	20.62
Total Composite	76	100	23788	87.17	10.95	659.22	46.97	0.91	3.23	0.87	11.97

NOTE: For the Kindergarten level the reliability and the SEM of the Prewriting subtest are missing because the teachers captured students' data at the subtest level rather than at the item level.

Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand
(continued)

3rd Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	23516	15.12	3.00	656.67	42.60	0.70	1.64	0.76	21.74
Writing	20	20	23516	15.61	3.71	666.69	52.25	0.81	1.60	0.78	20.55
Reading	20	20	23516	12.44	4.16	661.74	45.63	0.80	1.85	0.62	19.01
Speaking	14	32	23516	28.76	4.38	681.26	56.78	0.89	1.45	0.90	19.15
Comprehension	40	40	23516	27.57	6.45	658.38	38.12	0.85	2.50	0.69	14.42
Oral	34	52	23516	43.89	6.29	661.93	40.57	0.87	2.31	0.84	14.63
Total Writing	22	28	23516	20.68	4.62	661.95	49.50	0.65	2.72	0.74	19.01
Total Composite	76	100	23516	77.02	13.00	659.19	35.98	0.89	4.24	0.77	10.12

4th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	18850	16.11	2.66	670.90	42.01	0.68	1.52	0.81	21.74
Writing	20	20	18850	16.71	3.22	683.08	50.80	0.79	1.46	0.84	20.55
Reading	20	20	18850	13.85	4.01	677.56	46.86	0.80	1.77	0.69	19.01
Speaking	14	32	18850	29.31	4.00	689.06	54.89	0.89	1.34	0.92	21.67
Comprehension	40	40	18850	29.96	6.01	673.03	38.38	0.85	2.36	0.75	14.42
Oral	34	52	18850	45.42	5.68	673.64	41.64	0.86	2.13	0.87	14.63
Total Writing	22	28	18850	22.31	4.10	681.50	52.17	0.65	2.41	0.80	19.01
Total Composite	76	100	18850	81.58	11.78	673.28	36.37	0.89	3.86	0.82	10.5

5th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	15237	16.61	2.52	679.60	43.16	0.67	1.45	0.83	21.74
Writing	20	20	15237	17.25	2.95	692.26	49.94	0.79	1.37	0.86	22.75
Reading	20	20	15237	14.73	3.83	688.18	47.38	0.80	1.71	0.74	19.01
Speaking	14	32	15237	29.48	4.09	692.93	54.96	0.90	1.28	0.92	21.67
Comprehension	40	40	15237	31.34	5.71	682.41	38.93	0.84	2.26	0.78	14.95
Oral	34	52	15237	46.09	5.71	680.71	43.90	0.87	2.05	0.89	15.44
Total Writing	22	28	15237	23.13	3.83	693.73	54.77	0.62	2.35	0.83	20.51
Total Composite	76	100	15237	83.95	11.46	682.06	37.59	0.89	3.76	0.84	10.75

Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand
(continued)

6th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	12315	16.15	2.92	696.16	43.24	0.70	1.58	0.81	19.53
Writing	24	24	12315	20.52	3.60	711.29	50.95	0.82	1.53	0.85	18.62
Reading	24	24	12315	19.13	4.14	707.52	48.19	0.83	1.70	0.80	18.45
Speaking	14	32	12315	28.83	4.73	732.74	57.89	0.91	1.42	0.90	14.14
Comprehension	44	44	12315	35.28	6.41	699.95	40.27	0.87	2.34	0.80	13.83
Oral	34	52	12315	44.98	6.59	708.11	42.38	0.88	2.27	0.86	11.83
Total Writing	26	32	12315	26.29	4.42	717.44	55.98	0.72	2.33	0.82	17.22
Total Composite	84	108	12315	90.40	13.17	706.05	38.74	0.92	3.80	0.84	8.75

7th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	10474	16.31	3.02	699.57	45.38	0.74	1.55	0.82	19.53
Writing	24	24	10474	20.72	3.64	715.39	52.06	0.83	1.49	0.86	20.09
Reading	24	24	10474	19.60	4.13	714.61	50.31	0.84	1.64	0.82	18.45
Speaking	14	32	10474	28.78	5.12	733.87	59.82	0.92	1.41	0.90	14.70
Comprehension	44	44	10474	35.91	6.51	705.09	42.33	0.88	2.28	0.82	14.39
Oral	34	52	10474	45.09	7.20	710.77	45.39	0.90	2.26	0.87	12.11
Total Writing	26	32	10474	26.66	4.49	724.43	59.39	0.74	2.29	0.83	17.22
Total Composite	84	108	10474	91.35	13.94	710.97	42.16	0.93	3.74	0.85	9.17

8th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	8908	16.61	2.99	704.73	46.30	0.75	1.50	0.83	23.10
Writing	24	24	8908	21.10	3.44	721.67	51.47	0.83	1.42	0.88	20.09
Reading	24	24	8908	20.37	3.78	725.29	50.08	0.84	1.53	0.85	19.99
Speaking	14	32	8908	28.78	5.16	734.80	60.42	0.93	1.40	0.90	15.40
Comprehension	44	44	8908	36.98	6.20	713.21	43.20	0.88	2.15	0.84	15.12
Oral	34	52	8908	45.39	7.27	714.26	47.06	0.91	2.21	0.87	12.92
Total Writing	26	32	8908	27.27	4.31	734.77	61.51	0.84	1.71	0.85	18.20
Total Composite	84	108	8908	93.03	13.68	718.04	43.89	0.94	3.25	0.86	9.70

Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand
(continued)

9th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	8660	16.14	3.51	710.12	51.17	0.81	1.54	0.81	23.07
Writing	24	24	8660	18.86	3.87	709.00	43.47	0.79	1.76	0.79	17.78
Reading	24	24	8660	17.67	4.47	704.57	46.43	0.83	1.86	0.74	18.52
Speaking	14	32	8660	27.90	6.65	746.57	66.14	0.94	1.61	0.87	14.04
Comprehension	44	44	8660	33.81	7.33	704.87	44.05	0.89	2.43	0.77	14.46
Oral	34	52	8660	44.03	9.43	725.01	54.00	0.93	2.44	0.85	12.11
Total Writing	26	32	8660	24.75	4.89	706.67	40.96	0.72	2.60	0.77	17.54
Total Composite	84	108	8660	86.45	16.92	709.37	39.48	0.94	4.18	0.80	9.21

10th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	6432	16.28	3.37	711.89	50.14	0.80	1.52	0.81	23.07
Writing	24	24	6432	19.05	3.75	711.11	43.10	0.78	1.74	0.79	18.73
Reading	24	24	6432	18.20	4.28	709.89	45.79	0.82	1.82	0.76	18.52
Speaking	14	32	6432	28.02	6.14	745.69	63.64	0.93	1.62	0.88	14.04
Comprehension	44	44	6432	34.48	7.03	708.62	43.21	0.88	2.39	0.78	15.16
Oral	34	52	6432	44.30	8.69	725.13	51.51	0.92	2.42	0.85	12.57
Total Writing	26	32	6432	25.07	4.72	709.60	41.03	0.64	2.82	0.78	18.69
Total Composite	84	108	6432	87.57	15.85	711.83	38.19	0.92	4.48	0.81	9.38

11th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	5505	16.61	3.23	717.07	49.86	0.79	1.48	0.83	23.07
Writing	24	24	5505	19.49	3.51	715.95	41.94	0.77	1.70	0.81	20.16
Reading	24	24	5505	18.80	4.15	716.93	46.67	0.82	1.76	0.78	20.02
Speaking	14	32	5505	28.48	5.76	752.23	62.51	0.93	1.54	0.89	14.81
Comprehension	44	44	5505	35.42	6.81	714.86	43.72	0.88	2.32	0.80	15.16
Oral	34	52	5505	45.10	8.19	731.37	51.87	0.92	2.31	0.87	13.09
Total Writing	26	32	5505	25.72	4.45	715.50	40.73	0.70	2.43	0.80	18.69
Total Composite	84	108	5505	89.62	15.10	717.75	38.44	0.93	3.93	0.83	10.01

Table 4.1: Descriptive Statistics and Reliability in Raw and Scaled Score by Grade by Strand
(continued)

12th Grade											
Subtest	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	4816	16.75	2.99	718.49	48.15	0.76	1.47	0.84	27.06
Writing	24	24	4816	19.54	3.38	716.22	40.96	0.75	1.70	0.81	22.47
Reading	24	24	4816	19.15	3.86	720.38	45.08	0.80	1.74	0.80	20.02
Speaking	14	32	4816	28.73	5.28	754.07	60.12	0.92	1.48	0.90	15.75
Comprehension	44	44	4816	35.91	6.25	717.48	42.27	0.87	2.29	0.82	16.07
Oral	34	52	4816	45.49	7.41	732.36	49.32	0.91	2.25	0.87	13.72
Total Writing	26	32	4816	25.82	4.22	716.13	39.84	0.65	2.48	0.81	18.69
Total Composite	84	108	4816	90.46	13.69	719.22	36.32	0.92	3.91	0.84	10.01

4.5 Reliability of Classification Decision at Proficient Cut

Based on the AZELLA Form AZ-2 scaled scores, student performance is classified into one of five proficiency levels. While it is always important to know the reliability of student scores in any examination, it is of even greater importance to assess the reliability of the decisions based on these scores. Evaluation of the reliability of classification decisions is performed through estimation of the probabilities of correct and consistent classification of student performance. Procedures from Livingston and Lewis (1995) were applied to derive measures of the accuracy and consistency of the classifications. Brief descriptions of the procedures used and results obtained are presented in this section.

The accuracy of decisions is the extent to which decisions would agree with those that would be made if each student could somehow be tested with all possible forms of the examination. The consistency of decisions is the extent to which decisions would agree with the decisions that would have been made if the students had taken a parallel form of the AZELLA Form AZ-2 equal in difficulty and covering the same content as the form they actually took. These ideas are shown schematically in Figures 4.1 and 4.2. Please note that the term “Achieves Proficient Status” refers to the proficient category on the Total Composite score and “Does Not Achieve Proficient Status” refers to all categories below proficient status.

Figure 4.1: Classification Accuracy

Decision made on a form actually taken			
Does Not Achieve Proficient Status			Achieves Proficient Status
True status made on all-forms average	Does Not Achieve Proficient Status	Correct Classification	Misclassification
	Achieves Proficient Status	Misclassification	Correct Classification

Figure 4.2: Classification Consistency

Decision made on the 2nd form taken			
Does Not Achieve Proficient Status			Achieves Proficient Status
Decision made on the first form taken	Does Not Achieve Proficient Status	Correct Classification	Misclassification
	Achieves Proficient Status	Misclassification	Correct Classification

Note that Figures 4.1 and 4.2 were adapted from Young and Yoon (1998).

In Figure 4.1, accurate classifications occur when the decision made on the basis of the all-forms average (or true score) agrees with the decision made on the basis of the form actually taken.

Misclassifications occur when, for example, a student who actually accomplished “Does Not Achieve Proficient Status” on the basis of his or her all-forms average is classified incorrectly as accomplishing “Achieves Proficient Status.” Consistent classification occurs (Figure 4.2) when two forms agree on the classification of a student as either “Achieves Proficient Status” or “Does Not Achieve Proficient Status,” whereas inconsistent classification occurs when the decisions made by the forms differ.

These analyses make use of the techniques outlined and implemented by Hanson (1991), Haertel (1996), Livingston and Lewis (1995), and Young and Yoon (1998). The software *USmooth* developed by Hanson (1995) was used for the analyses. Estimates of decision accuracy and consistency were made for the “Achieves Proficient Status” cut points on the Total Composite score reported in the AZELLA Form AZ-2.

The table also includes the proportions of False Positive and False Negative classifications. The sum of values of Accuracy, False Positive, and False Negative is equal to 1.00, but due to rounding the table values may or may not equal 1.00. False Positive and False Negative classifications refer to the mismatch between student true scores and observed scores. The False Positive value is the proportion of student scores misclassified to the category “Achieves Proficient Status” when student scores do not meet proficient status. The False Negative value is the proportion of student scores misclassified to the category “Does Not Achieve Proficient Status” when student scores actually do meet proficient status.

Table 4.2 presents the results of the decision accuracy and consistency of the “Achieves Proficient Status” cut scores for the Total Composite score based on the 2009-2010 operational administration excluding students with missing subtest scores. The table contains the following:

- Accurate classifications;
- False positives;
- False negatives;
- Consistent classifications.

The table illustrates the general rule that decision consistency will be less than the decision accuracy. It should also be noted that the decision accuracy of students who achieved Proficient Status for the Total Test ranged from 0.89 to 0.96.

Table 4.2: Decision Accuracy and Consistency by Grade

Grade	Accuracy	False Positives	False Negatives	Consistency
K	0.90	0.10	0.00	0.86
1	0.91	0.05	0.04	0.87
2	0.96	0.02	0.02	0.94
3	0.89	0.06	0.05	0.85
4	0.89	0.05	0.06	0.85
5	0.91	0.04	0.05	0.86
6	0.95	0.02	0.03	0.93
7	0.95	0.02	0.03	0.93
8	0.95	0.02	0.03	0.94
9	0.93	0.03	0.04	0.90
10	0.92	0.04	0.05	0.88
11	0.92	0.03	0.04	0.89
12	0.91	0.04	0.05	0.87

5. VALIDITY

For the AZELLA Form AZ-2, Form C of the SELP was used together with augmented items created by Arizona teachers to construct one form per grade span. The AZELLA Form AZ-2 supports the validity-related standards set forth in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999). Our judgments about test validity are based on the following sources of evidence of validity:

- Test content—“an analysis of the relationship between a test’s content and the construct it is intended to measure” (p. 11)
- Internal structure—“the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are made” (p. 13)
- Relationships to other variables—“analyses of the relationship of test scores to variables external to the test” (p. 13)

5.1 Test Content

Evidence of validity based on test content is revealed by the extent to which the material on the test represents an appropriate sampling of skills, knowledge, and understanding of the domain tested. As part of the development of the AZELLA Form AZ-2, item writers were trained to write items representative of the intent of the Arizona K–12 English Language Learner Proficiency standards’ performance indicators. A critical part of the item review process included the appropriateness of the match of the item to the performance indicator being assessed. The items in the Pearson ELL item bank were reviewed to determine their match to individual performance indicators of Arizona ELL proficiency standards for each grade span. Only those items relating specifically to an instructional standard were included in the test forms. In addition, new items were created to better align the test to the standards.

5.2 Internal Structure

Because an English language proficiency test should be able to detect performance and proficiency differences among students, it is important to examine how well each item functions consistently with the overall intent of the test. Biserial correlation coefficients reveal how well an item discriminates between high- and low-scoring students. As test forms were developed, the fit of the construct being assessed was examined in terms of the way it was assessed and the way students were able to respond. Content experts were asked to examine the test blueprints and items to be sure that the test would logically relate to the most current empirical and theoretical understanding of the constructs being assessed.

5.3 Evidence of the Internal Structure of AZELLA

An assessment procedure should not be a random collection of assessment tasks or test questions. Each task in the assessment should contribute positively to the total result. The interrelationship among the tasks on an assessment is known as the internal structure of the assessment. Typical questions that investigate the relationships among assessment parts include (Nitko, 2004):

- Do all of the assessment tasks work together so that each task contributes positively toward assessing the quality of interest?
- If different parts of the assessment procedure are intended to provide unique information, do the results support this uniqueness?
- If different parts of the assessment procedure are intended to provide the same or similar information, do the results support this?

Correlations were obtained between the four subtests, Listening, Reading, Speaking, and Writing, in order to answer these questions. In theory, the relationship between the Reading and Writing subtests should be strong. Table 5.1 presents the intercorrelations among the four subtests by grade based on the 2009-2010 operational administration with students who took all subtests. The results of this analysis showed that the correlations between the Reading and Writing subtests ranged from 0.69 to 0.78. The correlations between the Reading and Writing subtests were consistently the highest compared to the other combinations of subtests for each grade.

The evidence of internal structure of the AZELLA Form AZ-2 is also depicted by the point biserial correlation coefficient and fit statistics. Appendices A.1–A.5 provide these statistics for the AZELLA Form AZ-2.

Table 5.1: Intercorrelations Among the Subtests by Grade

Kindergarten				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.60	1.00		
Speaking	0.59	0.53	1.00	
Total Writing	0.58	0.70	0.54	1.00

1st Grade				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.45	1.00		
Speaking	0.41	0.34	1.00	
Total Writing	0.46	0.77	0.36	1.00

2nd Grade				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.47	1.00		
Speaking	0.41	0.39	1.00	
Total Writing	0.49	0.78	0.42	1.00

3rd Grade				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.62	1.00		
Speaking	0.43	0.38	1.00	
Total Writing	0.63	0.71	0.41	1.00

4th Grade				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.61	1.00		
Speaking	0.43	0.35	1.00	
Total Writing	0.61	0.70	0.40	1.00

5th Grade				
	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.60	1.00		
Speaking	0.46	0.38	1.00	
Total Writing	0.60	0.69	0.44	1.00

6th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.64	1.00		
Speaking	0.46	0.40	1.00	
Total Writing	0.63	0.75	0.45	1.00

7th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.65	1.00		
Speaking	0.53	0.45	1.00	
Total Writing	0.66	0.76	0.51	1.00

8th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.68	1.00		
Speaking	0.56	0.48	1.00	
Total Writing	0.68	0.77	0.55	1.00

9th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.69	1.00		
Speaking	0.69	0.52	1.00	
Total Writing	0.75	0.75	0.66	1.00

10th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.69	1.00		
Speaking	0.64	0.49	1.00	
Total Writing	0.74	0.75	0.61	1.00

11th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.69	1.00		
Speaking	0.63	0.52	1.00	
Total Writing	0.74	0.75	0.61	1.00

12th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00			
Reading	0.66	1.00		
Speaking	0.57	0.46	1.00	
Total Writing	0.72	0.71	0.55	1.00

5.4 Evidence of the Unidimensionality of the AZELLA

A study was conducted to determine the dimensionality of the AZELLA. The Standards for Educational and Psychological Testing (AERA, APA, and NCME, 1999) define the internal structure of a test as “the degree to which the relationships among test items and components conform to the construct on which the proposed test score interpretations are made” (p. 13). The purpose of the AZELLA is to measure students’ proficiency in English, and one of the fundamental assumptions about the Rasch model used is that the test is unidimensional. If the AZELLA really measures a single construct, then the internal structure of the AZELLA should reflect only one dominant construct in the test.

According to Hattie (1985), a variety of methods are available for determining unidimensionality, and many different indices are used. These indices are based on one of the following: answer patterns, reliability, components and factor analysis, and latent traits. When defining unidimensionality based on principal components, some estimation issues arise. The issues include how to determine the number of factors, the problem of communalities, the role of eigenvalues, and the choice of correlations. Traditionally, the first principal component has been used as an index of unidimensionality. However, the obvious problem is how “high” the variance has to be in order to conclude that the test is unidimensional.

Lord’s (1980) suggested procedure for determining unidimensionality is to take the ratio of the first eigenvalue to the second eigenvalue and to verify that the second eigenvalue is not much larger than any of the others. Hattie (1985) suggested a possible index to operationalize Lord’s criteria by using the difference of eigenvalues between the first factor and the second factor divided by the difference of eigenvalues between the second factor and the third to evaluate unidimensionality. If the ratio is large (usually larger than 3), the first factor is relatively strong.

The AZELLA Form AZ-2 2009-2010 operational administration data with students who took all subtests was used to conduct the analyses. The analyses were carried out by test levels—Preliteracy, Primary, Elementary, Middle Grades, and High School. It must be noted that for the Preliteracy level the Prewriting subtest was not included in the analysis because student’s score is captured at the subtest-level and not at the item-level. The Principal Components Analysis (PCA) method was used to extract total variance, and the results are shown in Table 5.2. The ratios for all levels are larger than 3. The results suggest that there is one dominant factor for the AZELLA Form AZ-2.

Table 5.2: Results of Principal Components Analysis

Level	Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Ratio
Preliteracy	14.11	3.02	1.92	10.15
Primary	25.63	6.00	3.96	9.62
Elementary	23.11	5.40	2.78	6.75
Middle	29.73	5.83	1.95	6.15
High	30.86	4.63	2.07	10.22

5.5 Relationships to Other Variables

For the AZELLA Form AZ-2, evidence of validity based on relationships to other variables will be conducted as ongoing research. Data from the 2010 administration of the AIMS will be obtained in order to examine the relationship between English language proficiency as measured by AZELLA Form AZ-2 to the student achievement as measured by AIMS. This analysis was performed with data from AZELLA Form AZ-1 and is shown below.

The following table examines the academic achievement of former English language learners who were labeled “Proficient” based on their AZELLA AZ-1 test, to the achievement of all students in Arizona. Academic achievement was measured by student proficiency on the state’s Arizona Instrument to Measure Standards (AIMS) test for mathematics, reading, and writing. Students included in the analysis were all those who took the AIMS or the alternate assessment for the significantly cognitively disabled, the AIMS-A, from 2006 through 2009. Only test records with valid scores (that is excluding incomplete tests and tests taken with non-standard modifications) and valid student ID numbers were included. The “Reclassified” category included students who had been exited from the English language learner program based on AZELLA test scores for at least two years. Students were identified as exited if a reclassification transaction was submitted for the student to the state’s student information system two or more years in the past. The “All Students” category consisted of all students including English language learners and reclassified students.

The table shows that the reclassified students pass the AIMS assessments in mathematics, reading, and writing at rates equal to or higher than students in general in grades 3 through 7. Reclassified students in grades 8 and 10 tend to pass AIMS at rates within 2 to 7 percentage points of the students in general. These results provide evidence that reclassified students (i.e., students who had exited ELL programs based on their AZELLA scores), especially at the lower grades, are able to successfully engage in regular classroom studies as measured by their success on the state’s accountability measure.

Table 5.3: Statewide Comparisons of the Percent of Students Passing AIMS

Year/ Grade	Mathematics		Reading		Writing	
	Reclassified for 2 or More Years	All Students	Reclassified for 2 or More Years	All Students	Reclassified for 2 or More Years	All Students
2006						
3	75%	72%	69%	67%	56%	52%
4	86%	73%	81%	65%	79%	66%
5	81%	68%	79%	67%	71%	56%
6	71%	62%	74%	65%	96%	87%
7	74%	67%	71%	65%	95%	89%
8	55%	60%	58%	63%	85%	78%
10	62%	66%	67%	72%	59%	65%
2007						
3	88%	72%	85%	69%	92%	81%
4	85%	74%	75%	65%	89%	79%
5	82%	69%	80%	70%	80%	68%
6	71%	65%	70%	67%	87%	79%
7	73%	69%	68%	67%	88%	82%
8	59%	61%	59%	63%	76%	73%
10	62%	68%	68%	73%	71%	74%
2008						
3	84%	71%	82%	68%	87%	76%
4	89%	74%	85%	69%	82%	70%
5	80%	69%	78%	68%	73%	65%
6	74%	67%	73%	68%	79%	71%
7	75%	71%	73%	70%	69%	65%
8	59%	61%	66%	67%	79%	76%
10	63%	69%	67%	74%	64%	69%
2009						
3	92%	73%	93%	71%	93%	79%
4	89%	74%	87%	72%	93%	83%
5	87%	72%	88%	73%	90%	79%
6	76%	68%	76%	70%	93%	86%
7	80%	73%	77%	73%	88%	83%
8	63%	63%	70%	69%	82%	86%
10	68%	70%	71%	75%	69%	72%

6. CALIBRATION, EQUATING, AND SCALING

This section of the technical manual describes the calibration, equating, and scaling procedures conducted on the Forms Field Test Fall 2008 dataset. The items on the AZELLA Form AZ-2 were analyzed within the framework of item response theory (IRT). The Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model (PCM) (Masters, 1982) for polytomous items were used to develop, calibrate, equate, and scale the AZELLA Form AZ-2. These measurement models are regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All item and test analyses, including item-fit analysis, scaling, equating, diagnosis, and performance prediction, were accomplished within this framework. The statistical software used to calibrate and scale the AZELLA Form AZ-2 was *Winsteps* (Linacre & Wright, 2000).

6.1 The Rasch and Partial Credit Models

The most basic expression of the Rasch model is in the item characteristic curve (ICC). It shows the probability of a correct response to an item as a function of the ability level. The probability of a correct response is bounded by 1 (certainty of a correct response) and 0 (certainty of an incorrect response). The ability scale is, in theory, unbounded. In practice, the ability scale ranges from -4 to +4 logits (log-odds) for heterogeneous ability groups.

As an example, consider Figure 6.1, which depicts an item that falls at approximately 0.85 on the ability (horizontal) scale. When a person answers an item at the same level as his or her ability, then that person has a probability of roughly 50% of answering the item correctly. Another way of expressing this is that if there is a group of 100 people, all of whom have an ability of 0.85, it would be expected that about 50% of them would answer the item correctly. A person whose ability was above 0.85 would have a higher probability of getting the item right, while a person whose ability is below 0.85 would have a lower probability of getting the item right. This makes intuitive sense and is the basic formulation of Rasch measurement for test items having only two possible categories (i.e., wrong or right).

Figure 6.2 extends this formulation to show the probabilities of obtaining a wrong answer or a right answer. The curve on the left ($j=0$) shows the probability of getting a score of “0” (i.e., a wrong answer) while the curve on the right ($j=1$) shows the probability of getting a score of “1” (i.e., a correct answer). The point at which the two curves cross indicates the transition point on the ability scale where the most likely response changes from a “0” to a “1.” Here, the probability of answering the item correctly is 50%.

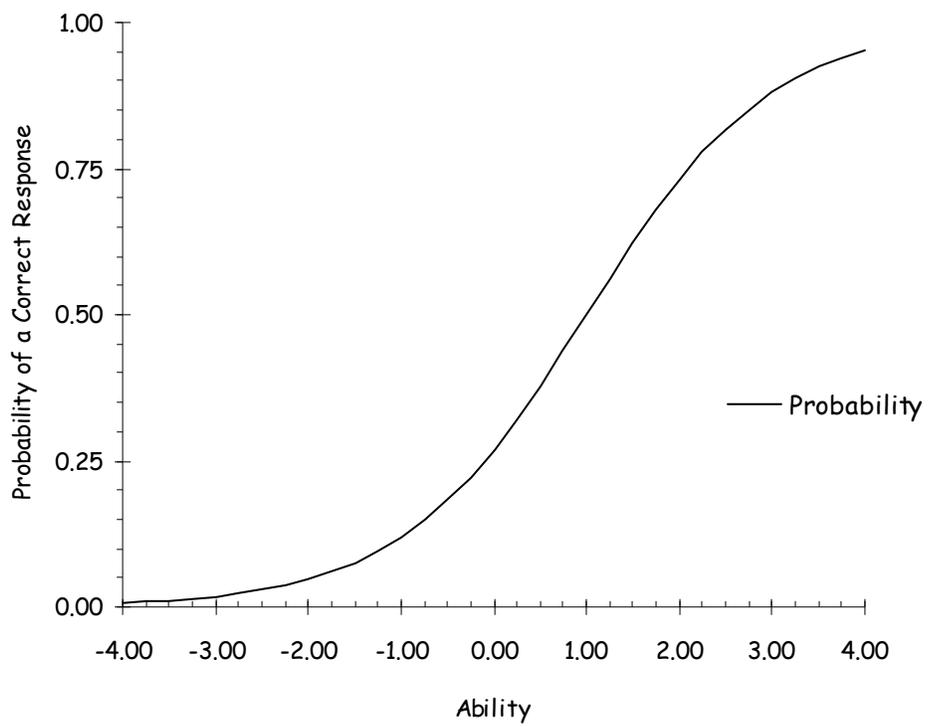


Figure 6.1: Sample Item Characteristic Curve

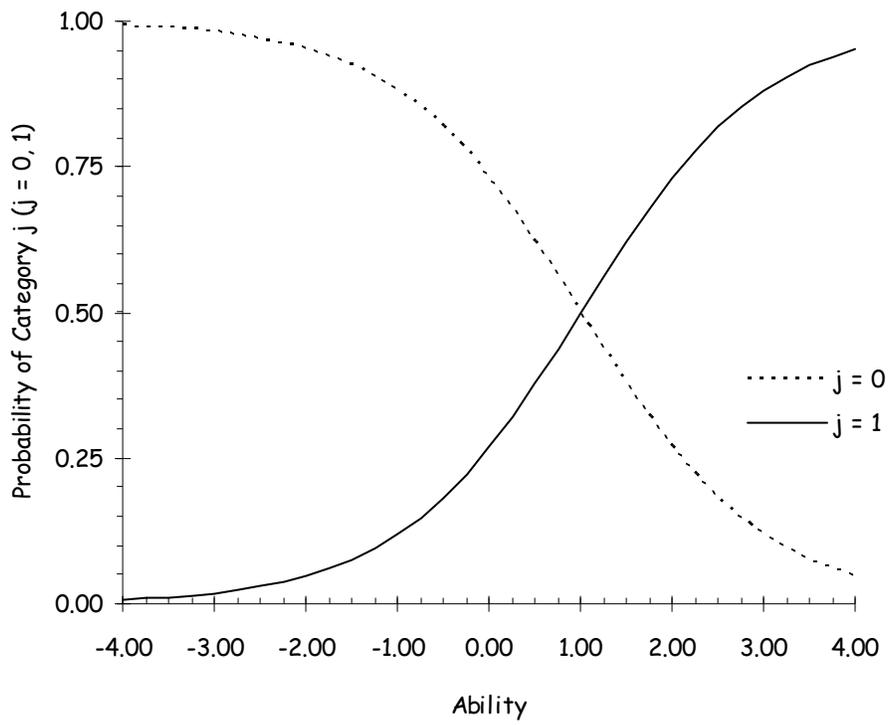


Figure 6.2: Category Response Curves for a Dichotomous or One-step Item

The key step in the formulation of the probabilities of obtaining a wrong answer or a correct answer and the point at which the Rasch dichotomous model merges with the polytomous Partial Credit Model requires us to assume an additional response category. Suppose that, rather than scoring items as completely wrong or completely right, a category is added to represent answers that, though not totally correct, are still clearly not totally incorrect. These relationships are shown in Figure 6.3.

The left-most curve ($j=0$) in Figure 6.3 represents the probability for all examinees getting a score of “0” (completely incorrect) on the item, given their ability. Those of very low ability (e.g., below -2) are very likely to be in this category and, in fact, are more likely to be in this category than the other two. Those receiving a “1” (partial credit) tend to fall in the middle range of abilities (the middle curve, $j=1$). The final, right-most curve ($j=2$), represents the probability for those receiving scores of “2” (completely correct). Very high-ability people are clearly more likely to be in this category than in any other, but there are still some of average and low ability that can get full credit for the item.

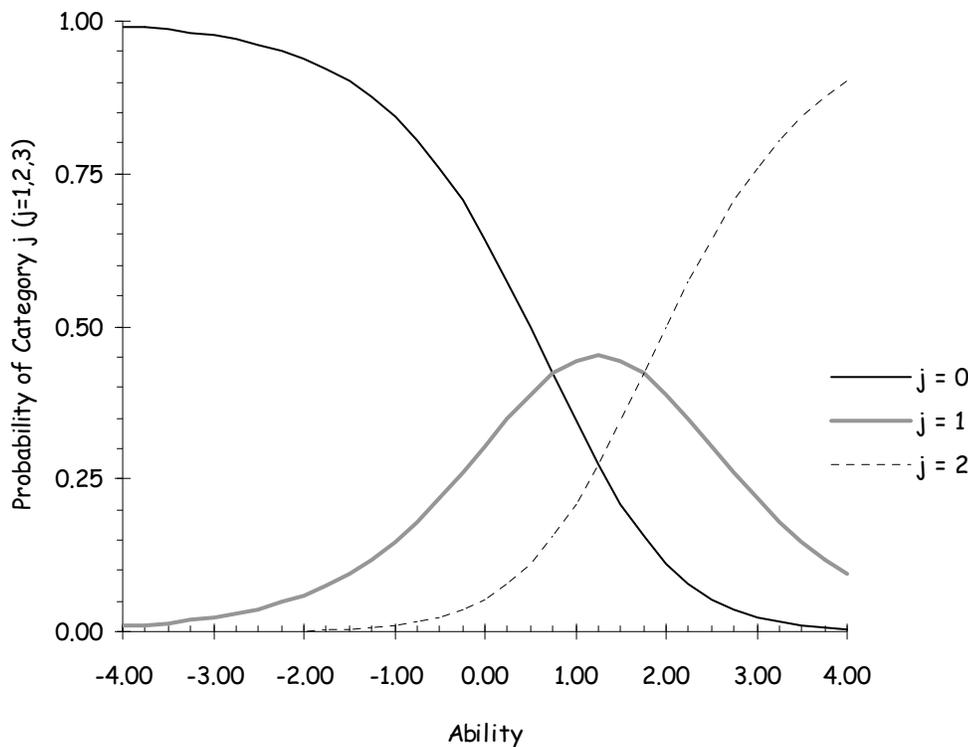


Figure 6.3: Category Response Curves for a Two-step Item

The actual computations for the PCM are quite complex; however, the points at which lines cross each other have a similar interpretation as the dichotomous case. Consider the point at which the $j=0$ line crosses the $j=1$ line, indicated by the left arrow. For abilities to the left of (or less than) this point, the probability is greatest for a “0” response. To the right of (or above) this point, and up to the point at which the $j=1$ and $j=2$ lines cross, the most likely response is a “1.” For abilities to the right of this point, the most likely response is a “2.”

Note that the probability of scoring a “1” response ($j=1$) declines in both directions, as ability decreases to the low extreme or increases to the high extreme. These points then may be thought of as the difficulties of crossing the thresholds between categories.

An important implication of the formulation can be summarized as: If the commonly used Rasch model applied to dichotomously (right/wrong) scored items can be thought of as simply a special case of the PCM, then the act of scaling multiple-choice items together with polytomous items, whether they have three or more response categories, is a straightforward process of applying the measurement model. The quality of the scaling then can be assessed in terms of known procedures.

One important property of the PCM is its ability to separate the estimation of item/task parameters from the person parameters. With the PCM, as with the Rasch model, the total score given by the sum of the categories in which a person responds is a sufficient statistic for estimating person ability (i.e., no additional information need be estimated). The total number of responses across examinees in a particular category is a sufficient statistic for estimating the step difficulty for that category. Thus with PCM, the same total score will yield the same ability estimate for different examinees.

The PCM is a direct extension of the dichotomous one-parameter IRT model developed by Rasch in the 1950s (Rasch, 1980). For an item/task involving m_i score categories, one general expression for the probability of scoring x on item/task i is given by

$$P_{xi} = \exp \sum_{j=0}^x (\theta - D_{ij}) / \sum_{k=0}^{m_i} \left[\exp \sum_{j=0}^k (\theta - D_{ij}) \right]$$

where $x = 0, 1, \dots, m_i$, and by definition,

$$\sum_{j=0}^0 (\theta - D_{ij}) = 0$$

The above equation gives the probability of scoring x on the i -th test item as a function of ability (θ) and the difficulty D_{ij} of the m_i steps of the task (Masters, 1982).

According to this model, the probability of an examinee scoring in a particular category (step) is the sum of the logit (log-odds) differences between θ and D_{ij} of all the completed steps, divided by the sum of the differences of all the steps of a task. Thissen and Steinberg (1983) refer to this model as a divide-by-total model. The parameters estimated by this model are (a) an ability estimate for each person (or ability estimate at each raw score level) and (b) m_i threshold (difficulty) estimates for each task with $m_i + 1$ score categories.

6.2 Calibration, Equating, and Scaling of the AZELLA Form AZ-2

A Forms Field Test of the AZELLA Form AZ-2 was conducted in the fall of 2008. The testing window was between October 14 and November 21, 2008. Calibration, equating, and scaling were conducted using data collected from this testing period. In the same way as was done for AZELLA Form AZ-1, Pearson used the pre-existing SELP vertical scale to create the vertical scale for Form AZ-2. SELP items, which comprised about 30% of the items on the AZELLA Form AZ-2, were fixed to the parameter values from the pre-existing vertical scale for the Preliteracy, Primary, Elementary, Middle Grades, and High School levels. That is, the SELP items were used as a common item link or anchor between the AZELLA Form AZ-2 and the

SELP item bank. Any remaining non-SELP items on the AZELLA Form AZ-2 were calibrated together with the SELP items using the Rasch and Partial Credit models. Fixing the values of the SELP items prior to calibration resulted in the item difficulty and step parameters of all the items being placed on the same ability metric as SELP and the AZELLA Form AZ-1.

Several iterations of *Winsteps* calibration were then run in order to determine the final sets of linking items to be used for the equating process. The following criteria were used:

- Rasch displacements with absolute values of equal to or greater than 0.5 are eliminated from the original linking items;
- At least 20% of the linking items have to be retained after the elimination process.

Table 6.1 below shows the number of original linking items and the total number of linking items after the iterations.

Table 6.1: Number of Linking Items

Level	# of Items	# Points	Links: Original	Links: Final
Preliteracy K	53	72	18	14
Primary 1–2	76	100	24	21
Elementary 3–5	76	100	23	20
Middle 6–8	84	108	28	22
High School 9–12	84	110	25	14

All test levels were able to pass these criteria except for the High School level. Here, the number of linking items fell below 20%. In this case, it was decided that while having the percentage of linking items fall below 20% was undesirable, it was more important to drop linking items that did not meet the Rasch displacement criterion.

After this process, the scale was then obtained by taking the item parameters for the Total Composite and using it to create raw score-to-scaled score tables. Subscales were also created for the reporting strands by taking the items for those strands and creating raw score-to-scale score tables. For example, the Reading subscale was created by using only the Reading items when creating the raw score-to-scale score table.

Finally, when these calibrations and scales were completed, the Forms Field Test Fall 2008 items were then calibrated to the pre-existing vertical scale. Appendices B.1–B.5 provide the raw score-to-scaled score conversion tables for the reporting strands by grade span.

6.4 Scaled Scores

The following equation was used to derive the AZELLA Form AZ-2 scaled scores:

$$SS = 35*\theta + 600$$

The AZELLA Form AZ-2 scaling procedure involves linear transformations of the Rasch ability (θ) into scaled score points (SS). These transformations do not give more weight to particular subtests, and they change neither the rank ordering of students nor their performance level classification.

7. IRT STATISTICS

7.1 Model and Rationale for Use

In addition to reporting raw score summary statistics and item level statistics using the classical test theory (CTT), Pearson also analyzed the items on the AZELLA Form AZ-2 test within the framework of Item Response Theory (IRT). The Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model (Masters, 1982) for polytomous items were used for developing, scoring, and reporting the AZELLA Form AZ-2 assessment. These models were used for several reasons.

First, the AZELLA vertical scale was created based on the pre-existing SELP vertical scale³ that was developed using the Rasch model. By using SELP items with known Rasch item difficulties, Pearson was able to create the AZELLA vertical scale.

Second, the sample size requirements for calibration, scaling, and equating under the Rasch and Partial Credit models are significantly smaller than for other IRT models. For example, the Rasch model requires on the order of 400 examinees per form for equating versus approximately 1,500 examinees per form under the 3PL IRT model (Kolen & Brennan, 2004, p. 288).

Finally, for the requirements of the AZELLA program, the Rasch model has a characteristic that makes it very useful—a one-to-one relationship between raw scores and scaled scores. That is, a student who answers a certain number of items correctly will receive the same scaled score as a second student with the same raw score, regardless of which particular items within the test form were answered correctly. These reasons lead Pearson to use the Rasch model as the IRT methodology for the AZELLA.

7.2 Evidence of Model Fit

Fit statistics are used for evaluating the goodness-of-fit of a model to the data. Fit statistics are calculated by comparing the observed and expected trace lines obtained for an item after parameter estimates are obtained using a particular model. *Winsteps* provides two kinds of fit statistics called mean-squares that show the size of the randomness or amount of distortion of the measurement system.

The OUTFIT and the INFIT statistics are used in order to ascertain the suitability of the data for constructing variables and making measures with the Rasch model. These fit statistics are mean square standardized residuals for item by person responses averaged over persons and partitioned between ability groups (OUTFIT) and within ability groups (INFIT). When the observed item characteristic curve (ICC) departs from the expected ICC from a reference value of 1, there is an expectation of high ability students failing on an easy item or low ability students succeeding on a difficult one. The OUTFIT mean square evaluates the agreement between the observed ICC and the best fitting Rasch model curve over the ability sub-groups. It is a standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's ability level. The INFIT, on the other hand, is a within-group mean square, which summarizes the degree of misfit remaining within ability groups after the between-group misfit has been removed from the total. The INFIT, therefore, is a standardized information-weighted mean square statistic, which is more sensitive to unexpected responses to items near the person's ability level.

OUTFIT mean-squares are influenced by outliers and are usually easy to diagnose and remedy. INFIT mean-squares, on the other hand, are influenced by response patterns and are harder to diagnose and remedy. In

³ The SELP vertical scale was created so that it was centered on the Elementary grades 3 – 5 span.

general, mean-squares near 1.0 indicate little distortion of the measurement system, while values less than 1.0 indicate that observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit).

Generally speaking, when item fit indices are lower than 0.5, they do not discriminate well and show greater than expected degree of consistency. Similarly, a fit value higher than 1.5 indicates inconsistency in examinee scores on the item (e.g., some unexpectedly high scores for low-ability candidates and low scores for high-ability candidates). There was 4% of the Middle Grades span items that were flagged for INFIT. None of the items in the other grade spans were flagged for INFIT. The percentage of items that was flagged for OUTFIT varied depending on grade spans. In general, around 1%–5% of items were flagged for OUTFIT.

The OUTFIT and the INFIT statistics are presented in the item statistics tables in Appendices A.1–A.5.

7.3 Rasch-Based Item Information

Table 7.1 presents the grade span, the subtest, the number of items in each subtest, the maximum number of points attainable for each subtest, and the average Rasch difficulty for each subtest based on the 2008 Forms Field Test data.

Table 7.1: Average Rasch Difficulty by Grade Span by Subtest

Level	Test	# of Items	Max Points	Average Rasch Difficulty
Preliteracy K	Listening	12	12	-4.37
	Prereading	12	12	-3.22
	Prewriting	17	22	-3.54
	Speaking	12	26	-3.46
	Total	53	72	-3.64
Primary 1–2	Listening	20	20	-2.64
	Writing Conventions	20	20	-0.63
	Reading	20	20	-0.65
	Speaking	14	32	-1.29
	Total	76	100	-1.26
Elementary 3–5	Listening	20	20	-0.20
	Writing Conventions	20	20	-0.07
	Reading	20	20	0.89
	Speaking	14	32	-0.86
	Total	76	100	0.02
Middle 6–8	Listening	20	20	0.66
	Writing Conventions	24	24	0.60
	Reading	24	24	1.13
	Speaking	14	32	0.50
	Total	84	108	0.77

Table 7.1: Average Rasch Difficulty by Grade Span by Subtest (Continued)

Level	Test	# of Items	Max Points	Average Rasch Difficulty
Level	Test	# of Items	Max Points	Average Rasch Difficulty
High School	Listening	20	20	1.03
9–12	Writing Conventions	24	24	1.19
	Reading	24	24	1.44
	Speaking	14	32	1.17
	Total	84	108	1.23

Appendices A.1–A.5 contain the results of the Forms Field Test Fall 2008 items of the AZELLA, which include the Rasch item parameters. The following IRT item parameters are presented for each item grouped by Listening/Speaking and Reading/Writing combinations:

- Item number
- Item format (multiple-choice, constructed-response, short-response, or extended-response)
- Maximum number of possible points
- N-Count (number of students)
- *P*-value for multiple-choice items (percentage of examinees that answered the item correctly)
- Item mean for constructed-response items (average number of points earned out of the maximum number of possible points)
- Point Biserial (index of discrimination between high- and low-scoring students)
- Vertically scaled Rasch item difficulty
- Standard Error of Rasch difficulty
- INFIT: Standardized information-weighted mean square statistic, which is sensitive to unexpected behavior affecting responses to items near the person’s ability level
- OUTFIT: Standardized outlier-sensitive mean square fit statistic that is sensitive to unexpected behavior by persons on items far from the person’s ability level.

8. STANDARD SETTING

8.1 Introduction

Standard setting was held for AZELLA Form AZ-1. As described in Chapter 6.2, AZELLA Form AZ-2 is on the same SELP vertical scale as the AZELLA Form AZ-1. Therefore, the AZELLA Form AZ-2 cut points were equated to the original AZELLA cut points via the SELP vertical scale. There was not a separate standard setting meeting for AZELLA Form AZ-2. Following is the description of the standard setting meeting conducted on the AZELLA Form AZ-1.

As the contractor for the AZELLA, Harcourt Assessment organized a performance standard setting meeting. The standard setting meeting was held over a two-day period from June 6–7, 2006, in Phoenix, Arizona. The purpose of this meeting was to provide preliminary recommendations for the English language proficiency cut scores on the Total Composite Score of the AZELLA Form AZ-1. Sections 8.2 through 8.9 provide descriptions of the AZELLA Form AZ-1 standard setting process.

For each of the standard setting committees, there was one psychometrics staff member from Harcourt Assessment to facilitate the technical part of the standard setting. In addition, specialists from Harcourt Assessment and officials from the ADE were present to provide support during the standard setting sessions. Data analyses were conducted by the Harcourt Assessment Psychometrics and Research Services department after the standard setting was completed.

8.2 Proficiency Categories for AZELLA

For the AZELLA Total Composite Score, there are five performance levels. The performance levels are:

- Pre-Emergent,
- Emergent,
- Basic,
- Intermediate, and
- Proficient.

8.3 Composition of Standard Setting Committees

The training and experience of the standard setting judges help establish the validity of the judges' ratings (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Although different criteria may be applied in the selection of educators and community representatives, it was recommended that all the individuals selected be familiar with ESL and how it is delivered in the classroom. The selection of teachers recommended to serve on a committee was based on the following criteria:

- Grade-level expertise and experience in English as a Second Language and/or English Language Arts;
- Instructional/supervisory experience with limited English proficient students;
- Balanced regional representation.

The AZELLA panelists were recruited by the ADE to participate in the standard setting meeting. Five separate standard setting groups totaling 77 panelists were established. Harcourt Assessment recommended that the Preliteracy, Primary, Elementary, Middle Grades, and High School committees set recommended standards on one grade each, as indicated in Table 8.1. The table also shows the number of judges who participated in each of the groups. The committees' two-day agenda for working through the tests is provided in Appendix C.1. The five groups worked separately. The committee members were informed that the cut scores were only recommended cut scores and that they were not the final reported cuts.

Table 8.1: Panel Composition for Standard Setting Committees

Group	Level	Grade	Number of Judges
1	Preliteracy	K	15
2	Primary	2	15
3	Elementary	4	18
4	Middle Grades	7	15
5	High School	9	14

All of the panelists completed the Committee Application form from which demographic information were collected. Table 8.2 provides the gender of the panelist. Table 8.3 shows the ethnicity of the panelists. Table 8.4 shows the area where panelists teach or work. Table 8.5 shows the number of years of classroom teaching experience of panelists. Table 8.6 provides the certifications that panelists have. Table 8.7 provides the panelists' endorsements or specialization. (The total numbers for Table 8.6 and Table 8.7 are greater than 77 because panelists could choose multiple options).

Table 8.2: Gender of Panelist

Gender	Count	Percent
Female	65	84
Male	11	14
Missing/Blank	1	1
Total	77	100

Table 8.3: Ethnicity of Panelist

Ethnicity	Count	Percent
Asian	1	1
Hispanic or Latino	24	31
Native American	7	9
White	38	49
Black	3	4
Other	1	1
Missing/Blank	3	4
Total	77	100

Table 8.4: Area Where Panelist Teach/Work

Area	Count	Percent
Rural	31	40
Suburban	9	12
Suburban/Urban	2	3
Urban	25	32
Missing/Blank	10	13
Total	77	100

Table 8.5: Years of Classroom Teaching Experience of Panelist

Experience	Count	Percent
0-5 yrs.	14	18
6-10 yrs.	26	34
11-15 yrs.	15	19
16-20 yrs.	7	9
20+ yrs.	13	17
Missing/Blank	2	3
Total	77	100

Table 8.6: Certification of Panelist

Certification	Count
Elementary	59
Secondary	23
Administrative	2
Special Education	4
Other	1
Total	89

Table 8.7: Specialization of Panelist

Specialization	Count
English as a Second Language (ESL)	47
Early Childhood	14
Structured English Immersion (SEI)	19
Bilingual Education	13
Reading	9
Special Education	3
Other	10
Total	115

8.4 The Standard Setting Process

Harcourt Assessment recommended a modified-Angoff procedure (Angoff, 1984) for producing the suggested cut scores for the AZELLA Form AZ-1. The modified-Angoff procedure conducted in this standard setting is sometimes referred to as the ACT/NAGB standard setting process (Reckase, 2000). This procedure has a long and successful history in similar applications for both educational and professional certification assessments. The modified-Angoff procedure provides a systematic technique for eliciting judgments from panels of experts and quantifying the results of the judgments. This method has been applied successfully and it is a widely recognized method to use when multiple-choice test items and open-ended items are being used (Hambleton & Plake, 1995). Moreover, research has shown that the modified-Angoff method produces ratings with better reliability and smaller variability among the ratings of judges than other standard setting procedures (Andrew & Hecht, 1976; Brennan & Lockwood, 1980; Cross, et al., 1984; Poggio, Glasnapp, & Eros, 1981; Skakun & Kling, 1980; Cizek, G. J., 1996). This procedure represents an appropriate balance between statistical rigor and informed opinion.

The standard setting activity for five groups took approximately 15 hours spread across two days. Orientation and initial training took place on the first morning in a large group setting. This meeting was followed by three separate but concurrently run sessions, which contained multiple rounds of ratings, discussion, and feedback.

8.5 Introduction to the Process

The first activity during the standard setting meeting was an orientation of the committee members to the standard setting process. The orientation of judges to the procedures for establishing cut scores for each proficiency level on the Total Composite Score was an important step to ensure the smooth operation of the standard setting. It is likely that most of the panel members were unfamiliar with standard setting, so acquainting them with the expectations for their performance served to increase their comfort and effectiveness.

At the outset, judges were reminded that their task was to review the items for their respective tests and to estimate the minimal acceptable performance for students at each proficiency level on each item. The orientation concentrated on helping judges become familiar with two substantive aspects of the standard setting procedure. First, the judges were asked to estimate how students who are just at the threshold for each proficiency level should perform, rather than how they do or will perform. This important distinction was emphasized on numerous occasions.

Second, judges were assured that their ratings would remain confidential. The recommended cut points were based on the group's ratings, and individual ratings were not released. The judges were told to feel free to raise questions during the sessions. Although an important goal of the process was for judges to approach consensus or convergence in ratings, it was integral to the process for judges to maintain a rating that they personally believed was correct, whether or not it was consistent with ratings made by other judges. As Fitzpatrick (1989) noted, preserving the anonymity of judges may make it easier for them to revise an initial item rating after they have learned more about the item because the judges have not been publicly committed to their initial rating of the item. In contrast, encouraging judges to maintain their initial ratings, if they believe them to be appropriate, may be desirable if it enables judges to resist pressures from other panel members to conform. In fact, Fitzpatrick suggests that conformity due to social pressure is not desirable in standard setting. Items with disparate ratings should be discussed in order to educate the judges about other judges' rationales behind their ratings. Any effects of undue social pressure should be moderated by the group leader during the standard setting.

8.6 Independent Ratings of Each Item

At the beginning of the breakout session, the group facilitator led the panelists in reviewing the previously established definitions of performance levels (Appendix C.2). They then led the panelists in developing a shared concept of the threshold student at each proficiency level in their respective grade span. Committee members were then each given copies of their respective tests and worked individually to answer the items. Once all committee members completed the entire test, answer keys were provided and the judges scored their tests. Committee members were given sufficient time (approximately 60–90 minutes) to independently rate each item on the test (Round 1). They were encouraged to read each item, consider the skills being assessed and the importance of those skills, think of 100 threshold students (at each proficiency level), and record an estimate of how many, or what percentage, of those 100 threshold students (at each proficiency level) should correctly answer the item. For the multiple-choice items, panelists identified the percentage of threshold students they believed should be able to correctly answer each item. For the open-ended items, panelists identified the average number of rubric points they believed a student at each performance level should score.

Upon completion of the first round of ratings, all secure materials were collected and inventoried before committee members were dismissed from the meeting. During the evening, the individual ratings of the judges were aggregated by the Harcourt Assessment research analysts. Statistics for each judge and for the entire panel were also computed. To obtain an overall estimate of the cut point for each proficiency level from the total group of judges, the initial item ratings provided by the judges were treated as *p*-values (proportions) and summed across items by level. The result of this summation is a number-correct value for each judge. The number-correct value was then averaged across judges to obtain the judges' estimate of the cut point for each of the proficiency levels.

8.7 Provision and Discussion of Data

The following morning, the judges' rating sheets, test booklets, and handouts were returned to them. On an overhead projector, judges were shown the frequency distributions of their individual item ratings and cut scores, along with the average cut score arrived at by their group. Once discussion of the results of the initial ratings concluded, judges were asked to review the entire set of items that they rated in Round 1, to reconsider these ratings in light of the data they were shown, and to revise any of their ratings, if necessary.

The judges' focus was again directed toward thinking about 100 threshold students at each proficiency level and how they should perform on the items. The Round 2 ratings were collected and inventoried along with the secure materials. As with Round 1, the judges' Round 2 ratings were aggregated. Statistics for each judge and for the entire panel were also computed.

8.8 Adjustment of Judges' Cut Scores

The judges' rating sheets, test booklets, and handouts were returned to them. On an overhead projector, judges were shown the frequency distributions of their Round 2 individual item ratings and cut scores, along with the average Round 2 cut score. Judges had the opportunity to alter their estimates of the Round 2 cut point if they felt that their Round 2 cut point was a little too high or low. The judges did not rate individual items in Round 3. The Round 3 cut scores were then collected and tabulated.

This process was repeated for each of the levels so that there was data from each of the judges from their respective groups. The median cut scores of the panelists at each of the proficiency cuts were used as the recommended cuts. Appendix C.3 provides the summary statistics by round for each of the committees. The IQR denotes the interquartile range, which is defined as the difference between the third quartile and the first quartile.

8.9 Analyses After Standard Setting

The median scores from the standard setting committees were used as the recommended cuts. The cut scores were based on the total AZELLA score. After the standard setting meetings, Harcourt Assessment performed several post-standard setting analyses. The following analyses were carried out:

- The first step was to look up the equivalent scaled scores corresponding to the raw score cuts recommended by the committees.
- Graphs were plotted using the grades as the independent variable and scaled score as the dependent variable. The four cut points were plotted on the same graph to show that the cuts were monotonically increasing from the lower cuts to the higher cuts.
- Impact analysis was conducted on the Forms Field Test data. The median raw scores from the third round of ratings were used as the cuts. The percentage of students falling into each of the proficiency level was calculated for the grades where the standards were set.
- A comparison was carried out between the Forms Field Test impact data and the previous year's SELP data.
- Smoothing of the scaled cut scores across all 13 grades, K–12, was conducted to ensure that there were no reversals.

Tables 8.8a and 8.8b provide the approved AZELLA Form AZ-1 cut score ranges in scaled score for the reporting strands⁴ for all grades.

⁴ See Chapter 3.3 for definition of reporting strands.

Table 8.8a: Scaled Score Cut Ranges (Grades K–5)

SUBTEST/GRADE	K	1	2	3	4	5
Listening						
Pre-Emergent	300-394	300-432	300-446	300-531	300-531	300-540
Emergent	395-446	433-458	447-469	532-557	532-557	541-567
Basic	447-492	459-519	470-530	558-610	558-610	568-619
Intermediate	493-537	520-585	531-603	611-657	611-657	620-657
Proficient	538-900	586-900	604-900	658-900	658-900	658-900
Writing Conventions						
Pre-Emergent	-	300-517	300-524	300-528	300-534	300-539
Emergent	-	518-559	525-564	529-566	535-569	540-575
Basic	-	560-602	565-604	567-609	570-618	576-624
Intermediate	-	603-639	605-647	610-651	619-657	625-667
Proficient	-	640-900	648-900	652-900	658-900	668-900
Prereading/Reading						
Pre-Emergent	300-427	300-516	300-531	300-551	300-551	300-555
Emergent	428-474	517-542	532-553	552-579	552-579	556-590
Basic	475-517	543-595	554-604	580-636	580-636	591-644
Intermediate	518-561	596-641	605-654	637-684	637-684	645-684
Proficient	562-900	642-900	655-900	685-900	685-900	685-900
Speaking						
Pre-Emergent	300-440	300-512	300-519	300-534	300-543	300-550
Emergent	441-472	513-530	520-535	535-562	544-562	551-567
Basic	473-532	531-572	536-576	563-601	563-606	568-610
Intermediate	533-572	573-610	577-617	602-642	607-649	611-649
Proficient	573-900	611-900	618-900	643-900	650-900	650-900
Comprehension (Listening+Prereading/Reading)						
Pre-Emergent	300-426	300-469	300-477	300-540	300-547	300-547
Emergent	427-460	470-492	478-505	541-567	548-567	548-578
Basic	461-512	493-570	506-576	568-623	568-627	579-632
Intermediate	513-560	571-622	577-636	624-672	628-672	633-678
Proficient	561-900	623-900	637-900	673-900	673-900	679-900
Oral (Listening+Speaking)						
Pre-Emergent	300-434	300-479	300-485	300-533	300-544	300-548
Emergent	435-462	480-506	486-510	534-560	545-564	549-567
Basic	463-524	507-562	511-565	561-604	565-610	568-613
Intermediate	525-568	563-603	566-613	605-648	611-652	614-657
Proficient	569-900	604-900	614-900	649-900	653-900	658-900
Total Test						
Pre-Emergent	300-443	300-505	300-511	300-539	300-548	300-553
Emergent	444-469	506-529	512-536	540-563	549-567	554-573
Basic	470-532	530-587	537-589	564-614	568-620	574-622
Intermediate	533-589	588-636	590-645	615-663	621-668	623-674
Proficient	590-900	637-900	646-900	664-900	669-900	675-900

Total Writing (Writing Conventions+Writing)

Pre-Emergent	300-438	300-543	300-543	300-544	300-545	300-545
Emergent	439-468	544-561	544-562	545-563	546-564	546-568
Basic	469-554	562-616	563-616	564-617	565-618	569-618
Intermediate	555-621	617-656	617-659	618-664	619-664	619-675
Proficient	622-900	657-900	660-900	665-900	665-900	676-900

Table 8.8b: Scaled Score Cut Ranges (Grades 6–12)

SUBTEST/GRADE	6	7	8	9	10	11	12
Listening							
Pre-Emergent	300-541	300-541	300-556	300-574	300-574	300-574	300-589
Emergent	542-569	542-569	557-581	575-602	575-602	575-602	590-613
Basic	570-628	570-628	582-637	603-658	603-658	603-658	614-667
Intermediate	629-677	629-677	638-691	659-707	659-707	659-707	668-720
Proficient	678-900	678-900	692-900	708-900	708-900	708-900	721-900
Writing Conventions							
Pre-Emergent	300-544	300-549	300-559	300-561	300-566	300-572	300-584
Emergent	545-577	550-583	560-589	562-595	567-597	573-604	585-613
Basic	578-627	584-634	590-641	596-644	598-652	605-659	614-667
Intermediate	628-673	635-682	642-689	645-694	653-701	660-712	668-725
Proficient	674-900	683-900	690-900	695-900	702-900	713-900	726-900
Reading							
Pre-Emergent	300-556	300-571	300-571	300-577	300-577	300-591	300-591
Emergent	557-593	572-593	572-602	578-612	578-612	592-612	592-621
Basic	594-646	594-646	603-653	613-665	613-665	613-665	622-672
Intermediate	647-693	647-693	654-703	666-711	666-711	666-721	673-721
Proficient	694-900	694-900	704-900	712-900	712-900	722-900	722-900
Speaking							
Pre-Emergent	300-559	300-559	300-566	300-579	300-579	300-586	300-586
Emergent	560-576	560-581	567-585	580-597	580-597	587-602	587-607
Basic	577-615	582-619	586-622	598-638	598-642	603-642	608-646
Intermediate	616-656	620-663	623-670	639-680	643-680	643-686	647-694
Proficient	657-900	664-900	671-900	681-900	681-900	687-900	695-900
Comprehension (Listening+ Reading)							
Pre-Emergent	300-548	300-556	300-563	300-576	300-583	300-590	300-590
Emergent	549-582	557-587	564-592	577-607	584-607	591-612	591-617
Basic	583-638	588-642	593-646	608-662	608-666	613-666	618-670
Intermediate	639-686	643-692	647-698	663-709	667-715	667-715	671-721
Proficient	687-900	693-900	699-900	710-900	716-900	716-900	722-900
Oral (Listening+Speaking)							
Pre-Emergent	300-553	300-558	300-563	300-577	300-577	300-582	300-587
Emergent	554-574	559-578	564-584	578-599	578-602	583-605	588-609
Basic	575-619	579-624	585-627	600-645	603-647	606-650	610-653
Intermediate	620-664	625-668	628-678	646-690	648-694	651-699	654-704
Proficient	665-900	669-900	679-900	691-900	695-900	700-900	705-900
Total Test							
Pre-Emergent	300-556	300-561	300-568	300-580	300-582	300-587	300-592
Emergent	557-580	562-583	569-588	581-604	583-606	588-609	593-613
Basic	581-629	584-634	589-636	605-655	607-658	610-659	614-662
Intermediate	630-676	635-683	637-691	656-702	659-706	660-711	663-717
Proficient	677-900	684-900	692-900	703-900	707-900	712-900	718-900

Total Writing (Writing Conventions+Writing)

Pre-Emergent	300-546	300-552	300-561	300-565	300-576	300-576	300-586
Emergent	547-576	553-576	562-583	566-602	577-602	577-609	587-609
Basic	577-626	577-631	584-636	603-655	603-661	610-661	610-666
Intermediate	627-676	632-680	637-689	656-705	662-713	662-713	667-722
Proficient	677-900	681-900	690-900	706-900	714-900	714-900	723-900

9. AZELLA Form AZ-2 2009-2010 ADMINISTRATION RESULTS

This section presents the results of the AZELLA Form AZ-2 2009-2010 administration operational data.

9.1 Results Based on All Student Records

Students may have taken the AZELLA Form AZ-2 more than once during the 2009-2010 school year. All student records are included in the analyses. Analyses are provided for all the reporting strands. The following are the reporting strands:

- Listening,
- Writing Conventions,
- Reading,
- Speaking,
- Comprehension (Listening + Reading),
- Oral (Listening + Speaking),
- Total Writing (Writing Conventions + Open-Ended Writing), and
- Total Test (Listening + Writing Conventions + Open Ended Writing + Reading + Speaking).

Table 9.1 shows the percentages of students in each of the proficiency categories by grade. The table also provides the total N-counts corresponding to the proficiency categories.

Table 9.2 and Table 9.3 provide the raw score and scaled score descriptive statistics by grade. The tables include the following information:

- Number of students,
- Means,
- Medians,
- Inter-quartile range (IQR), and
- Standard deviations.

9.2 Results Without Reclassified English Language Learners Students

Recall that AZELLA is not only given to students in Structured English Immersion (SEI) programs, but also to students who have been reclassified as “fluent English proficient” for a period of up to two years. In order to see the results for students in SEI programs alone, the reclassified students need to be removed from the data before analysis.

Tables 9.4 and 9.5 present results based on student data where English language learners deemed proficient through assessment with AZELLA were reclassified and removed from the data. Table 9.4 shows this for the Total Test scores aggregated across the 2009-2010 school year while Table 9.5 shows the results for the Total Test scores broken out by time of year (i.e., Fall, Mid-year, or Spring).

The tables show:

- Grade;
- Time of year (Table 9.5 only);
- Number of items (Table 9.4 only);
- Maximum number of possible points (Table 9.4 only);
- Number of students;
- Means and standard deviations in raw scores
- Means and standard deviations in scaled scores;
- Cronbach's alpha internal consistency reliability estimate;
- Standard error of measurement;
- Mean raw score as a proportion of the maximum obtainable score;
- IRT based conditional standard errors of measurement for the Proficient cut scores.

Table 9.1: Percentage of Students in Each Proficiency Category

Proficiency Level Percentage							
Grade	Strand	N-Count	Pre-Emergent	Emergent	Basic	Intermediate	Proficient
K	Listening	45349	4	5	24	36	31
	Reading	45349	6	12	35	21	26
	Speaking	45349	14	7	35	20	25
	Comprehension	45349	5	8	38	29	21
	Oral	45349	9	9	35	27	20
	Total Writing	45349	23	8	35	20	14
	Total	45349	11	9	41	28	11
1	Listening	25855	1	0	4	25	70
	Writing Conventions	25855	3	8	25	29	35
	Reading	25855	2	2	34	33	29
	Speaking	25855	5	2	13	23	58
	Comprehension	25855	1	0	14	48	36
	Oral	25855	2	1	9	25	63
	Total Writing	25855	8	5	33	32	22
Total	25855	2	2	21	47	29	
2	Listening	24680	1	0	2	20	77
	Writing Conventions	24680	1	4	10	24	60
	Reading	24680	2	2	13	27	56
	Speaking	24680	3	1	6	20	70
	Comprehension	24680	1	0	5	28	66
	Oral	24680	2	1	4	20	74
	Total Writing	24680	3	2	12	25	58
Total	24680	2	1	7	31	59	
3	Listening	24314	2	1	12	35	50
	Writing Conventions	24314	2	3	13	23	59
	Reading	24314	3	2	25	43	27
	Speaking	24314	3	1	8	16	72
	Comprehension	24314	1	1	19	41	37
	Oral	24314	2	1	7	28	62
	Total Writing	24314	3	2	16	38	41
Total	24314	2	1	10	41	46	
4	Listening	19471	1	1	7	26	65
	Writing Conventions	19471	2	2	10	23	63
	Reading	19471	2	1	16	39	41
	Speaking	19471	3	0	6	19	72
	Comprehension	19471	1	1	13	31	53
	Oral	19471	3	1	5	24	67
	Total Writing	19471	2	1	9	30	58
Total	19471	2	1	7	34	56	
5	Listening	15674	2	1	7	19	72
	Writing Conventions	15674	2	2	7	18	71
	Reading	15674	2	2	14	31	51
	Speaking	15674	3	1	6	15	75
	Comprehension	15674	1	1	11	29	57
	Oral	15674	3	1	5	19	74

	Total Writing	15674	2	1	6	34	57
	Total	15674	2	1	6	31	60
6	Listening	12759	2	1	6	26	66
	Writing Conventions	12759	2	1	5	15	77
	Reading	12759	2	1	11	29	57
	Speaking	12759	3	0	2	8	86
	Comprehension	12759	2	1	7	27	64
	Oral	12759	2	1	2	10	85
	Total Writing	12759	2	1	5	15	78
	Total	12759	2	1	4	16	78
7	Listening	10963	2	1	7	23	68
	Writing Conventions	10963	2	1	7	18	73
	Reading	10963	2	1	10	25	62
	Speaking	10963	4	1	2	9	83
	Comprehension	10963	2	1	8	26	63
	Oral	10963	3	1	4	10	82
	Total Writing	10963	2	1	7	15	75
	Total	10963	2	1	5	17	75
8	Listening	9300	2	1	7	28	62
	Writing Conventions	9300	2	1	7	20	70
	Reading	9300	2	1	9	26	62
	Speaking	9300	4	1	3	11	81
	Comprehension	9300	2	1	8	24	66
	Oral	9300	3	1	4	15	77
	Total Writing	9300	2	1	6	16	75
	Total	9300	2	1	5	19	73
9	Listening	9611	6	4	13	38	40
	Writing Conventions	9611	4	2	11	24	59
	Reading	9611	5	3	18	36	38
	Speaking	9611	10	1	4	10	74
	Comprehension	9611	5	2	17	35	41
	Oral	9611	7	3	7	13	70
	Total Writing	9611	4	3	14	29	50
	Total	9611	5	3	10	25	56
10	Listening	6742	2	2	12	39	45
	Writing Conventions	6742	2	1	11	32	55
	Reading	6742	2	2	15	36	46
	Speaking	6742	5	1	4	11	78
	Comprehension	6742	2	1	17	39	41
	Oral	6742	3	2	6	18	71
	Total Writing	6742	2	1	14	37	46
	Total	6742	2	2	9	30	57
11	Listening	5689	1	2	10	37	50
	Writing Conventions	5689	1	1	10	40	48
	Reading	5689	1	1	12	43	43
	Speaking	5689	4	1	3	13	79
	Comprehension	5689	1	1	14	36	48
	Oral	5689	2	1	6	18	72
	Total Writing	5689	1	2	10	34	54
	Total	5689	1	1	8	31	59

12	Listening	4969	1	1	12	35	50
	Writing Conventions	4969	1	1	12	54	32
	Reading	4969	1	1	12	41	44
	Speaking	4969	3	1	4	15	78
	Comprehension	4969	1	1	12	44	42
	Oral	4969	1	1	6	21	71
	Total Writing	4969	1	1	12	46	41
	Total	4969	1	1	8	38	53

Table 9.2: Raw Score Descriptive Statistics by Grade

Grade	Strand	N-Count	Mean	Median	Std Dev	IQR
K	Listening	45349	9.00	10	2.51	3
	Reading	45349	7.62	8	3.36	6
	Speaking	45349	17.29	19	7.42	10
	Comprehension	45349	16.61	17	5.37	8
	Oral	45349	26.29	29	9.27	13
	Prewriting	45349	12.81	14	6.98	12
	Total	45349	46.72	50	17.29	26
1	Listening	25855	17.69	19	3.25	3
	Writing Conventions	25855	13.83	15	4.62	7
	Reading	25855	13.45	14	4.39	6
	Speaking	25855	24.59	27	7.42	8
	Comprehension	25855	31.13	32	6.82	8
	Oral	25855	42.28	45	9.75	10
	Total Writing	25855	18.14	19	6.16	9
	Total	25855	73.87	77	17.82	21
2	Listening	24680	18.74	20	2.66	1
	Writing Conventions	24680	16.79	18	3.90	3
	Reading	24680	16.62	18	3.82	4
	Speaking	24680	27.08	29	6.46	6
	Comprehension	24680	35.36	37	5.91	5
	Oral	24680	45.82	48	8.46	7
	Total Writing	24680	22.59	24	5.43	5
	Total	24680	85.02	90	15.97	14
3	Listening	24314	14.82	16	3.49	4
	Writing Conventions	24314	15.25	16	4.21	5
	Reading	24314	12.16	13	4.43	7
	Speaking	24314	28.03	30	6.14	5
	Comprehension	24314	26.98	28	7.25	11
	Oral	24314	42.85	45	8.67	7
	Total Writing	24314	20.18	22	5.41	6
	Total	24314	75.19	79	16.50	19
4	Listening	19471	15.79	17	3.29	3
	Writing Conventions	19471	16.34	18	3.88	4
	Reading	19471	13.54	15	4.39	6
	Speaking	19471	28.53	31	6.07	4
	Comprehension	19471	29.33	31	7.05	9
	Oral	19471	44.32	47	8.54	6
	Total Writing	19471	21.77	23	5.11	5
	Total	19471	79.63	84	16.14	15
5	Listening	15674	16.32	17	3.18	3
	Writing Conventions	15674	16.91	18	3.65	3
	Reading	15674	14.43	16	4.24	6
	Speaking	15674	28.81	31	5.91	3
	Comprehension	15674	30.75	33	6.80	8
	Oral	15674	45.12	48	8.33	6
	Total Writing	15674	22.65	24	4.86	5
	Total	15674	82.20	87	15.64	13

6	Listening	12759	15.77	17	3.65	4
	Writing Conventions	12759	20.03	21	4.60	4
	Reading	12759	18.65	20	4.97	5
	Speaking	12759	27.97	30	6.70	5
	Comprehension	12759	34.42	37	8.05	8
	Oral	12759	43.75	47	9.48	7
	Total Writing	12759	25.63	27	5.80	5
	Total	12759	88.03	94	18.56	14
7	Listening	10963	15.85	17	3.86	3
	Writing Conventions	10963	20.10	22	4.83	4
	Reading	10963	19.00	21	5.13	5
	Speaking	10963	27.65	31	7.43	5
	Comprehension	10963	34.84	38	8.44	8
	Oral	10963	43.50	47	10.47	6
	Total Writing	10963	25.82	28	6.16	5
	Total	10963	88.32	95	20.12	15
8	Listening	9300	16.16	17	3.83	4
	Writing Conventions	9300	20.54	22	4.59	3
	Reading	9300	19.81	21	4.82	4
	Speaking	9300	27.71	31	7.37	5
	Comprehension	9300	35.97	39	8.17	7
	Oral	9300	43.87	47	10.39	7
	Total Writing	9300	26.50	28	5.93	4
	Total	9300	90.17	97	19.70	12
9	Listening	9611	14.95	17	5.09	6
	Writing Conventions	9611	17.60	20	5.57	5
	Reading	9611	16.41	18	5.92	8
	Speaking	9611	25.33	30	10.13	8
	Comprehension	9611	31.37	35	10.53	12
	Oral	9611	40.28	47	14.64	12
	Total Writing	9611	22.95	26	7.41	8
	Total	9611	79.65	90	26.61	25
10	Listening	6742	15.80	17	4.13	5
	Writing Conventions	6742	18.53	20	4.59	5
	Reading	6742	17.67	19	5.00	6
	Speaking	6742	26.87	30	8.12	6
	Comprehension	6742	33.47	36	8.58	10
	Oral	6742	42.67	47	11.50	9
	Total Writing	6742	24.31	26	6.00	6
	Total	6742	84.65	92	20.90	19
11	Listening	5689	16.29	17	3.80	4
	Writing Conventions	5689	19.17	20	4.06	4
	Reading	5689	18.49	20	4.61	6
	Speaking	5689	27.66	31	7.35	5
	Comprehension	5689	34.78	37	7.87	8
	Oral	5689	43.95	48	10.36	8
	Total Writing	5689	25.24	27	5.32	6
	Total	5689	87.68	94	18.69	17

12	Listening	4969	16.54	18	3.40	4
	Writing Conventions	4969	19.34	20	3.71	4
	Reading	4969	18.95	20	4.18	5
	Speaking	4969	28.00	31	6.86	5
	Comprehension	4969	35.49	37	6.99	7
	Oral	4969	44.54	48	9.28	7
	Total Writing	4969	25.51	27	4.78	5
	Total	4969	89.00	94	16.38	16

Table 9.3: Scaled Score Descriptive Statistics by Grade

Grade	Strand	N-Count	Mean	Median	Std Dev	IQR
K	Listening	45349	513.02	527	61.40	75
	Reading	45349	515.65	512	65.17	99
	Speaking	45349	520.54	522	71.43	72
	Comprehension	45349	513.50	509	59.54	72
	Oral	45349	516.81	523	60.84	73
	Prewriting	45349	509.11	513	103.17	140
	Total	45349	514.15	515	61.46	75
1	Listening	25855	614.06	627	58.43	97
	Writing Conventions	25855	619.93	620	57.60	74
	Reading	25855	614.18	612	54.96	59
	Speaking	25855	620.14	621	70.08	72
	Comprehension	25855	612.41	608	55.12	57
	Oral	25855	617.07	615	61.38	66
	Total Writing	25855	619.13	617	59.83	72
	Total	25855	612.54	612	51.57	53
2	Listening	24680	639.29	678	52.19	51
	Writing Conventions	24680	661.37	661	58.47	56
	Reading	24680	659.03	662	59.06	68
	Speaking	24680	647.56	643	70.28	81
	Comprehension	24680	655.37	653	60.39	73
	Oral	24680	646.33	639	64.29	84
	Total Writing	24680	671.23	668	67.86	73
	Total	24680	652.89	655	59.16	65
3	Listening	24314	652.43	660	50.07	49
	Writing Conventions	24314	661.93	657	59.27	61
	Reading	24314	657.57	663	52.38	65
	Speaking	24314	674.11	671	70.56	113
	Comprehension	24314	654.14	657	46.55	59
	Oral	24314	656.54	658	51.69	48
	Total Writing	24314	656.95	664	57.28	59
	Total	24314	654.12	658	47.27	52
4	Listening	19471	666.39	675	50.76	46
	Writing Conventions	19471	677.98	688	59.46	69
	Reading	19471	672.94	682	54.85	60
	Speaking	19471	681.31	697	70.98	103
	Comprehension	19471	668.44	673	48.22	55
	Oral	19471	667.77	673	54.24	49
	Total Writing	19471	676.03	675	61.24	60
	Total	19471	667.73	673	49.52	49
5	Listening	15674	675.33	675	51.87	46
	Writing Conventions	15674	687.50	688	58.66	58
	Reading	15674	683.89	693	55.25	69
	Speaking	15674	686.17	697	69.50	92
	Comprehension	15674	678.11	686	48.60	54
	Oral	15674	675.41	683	55.39	61
	Total Writing	15674	688.59	688	63.52	76
	Total	15674	676.96	684	49.96	50

6	Listening	12759	691.10	697	53.19	53
	Writing Conventions	12759	705.54	700	61.34	67
	Reading	12759	702.03	703	57.98	58
	Speaking	12759	724.52	720	73.51	112
	Comprehension	12759	694.63	701	51.68	51
	Oral	12759	701.74	709	55.82	51
	Total Writing	12759	711.25	708	66.35	65
	Total	12759	699.84	707	52.98	47
7	Listening	10963	693.34	697	56.45	43
	Writing Conventions	10963	708.18	716	64.37	67
	Reading	10963	707.79	716	61.69	58
	Speaking	10963	723.05	746	79.04	112
	Comprehension	10963	698.49	708	55.27	51
	Oral	10963	702.70	709	60.38	46
	Total Writing	10963	716.56	722	71.48	83
	Total	10963	703.20	710	57.71	53
8	Listening	9300	698.54	697	57.37	71
	Writing Conventions	9300	715.07	716	62.92	56
	Reading	9300	718.91	716	61.12	67
	Speaking	9300	724.44	746	78.91	112
	Comprehension	9300	706.87	715	55.78	53
	Oral	9300	706.51	709	61.10	61
	Total Writing	9300	727.35	722	72.84	73
	Total	9300	710.64	718	58.37	49
9	Listening	9611	694.76	707	71.00	90
	Writing Conventions	9611	694.91	712	64.45	51
	Reading	9611	690.10	698	67.65	70
	Speaking	9611	722.69	736	97.31	132
	Comprehension	9611	689.80	701	67.31	63
	Oral	9611	706.14	722	79.36	81
	Total Writing	9611	691.88	709	63.63	63
	Total	9611	692.83	708	67.37	61
10	Listening	6742	705.62	707	59.43	80
	Writing Conventions	6742	705.16	712	53.73	58
	Reading	6742	703.86	707	56.17	56
	Speaking	6742	735.13	736	80.24	119
	Comprehension	6742	702.40	707	54.50	60
	Oral	6742	717.17	722	63.94	69
	Total Writing	6742	703.27	709	52.48	50
	Total	6742	704.78	714	52.57	50
11	Listening	5689	712.99	707	56.33	70
	Writing Conventions	5689	712.52	712	48.02	50
	Reading	5689	713.45	718	52.47	65
	Speaking	5689	744.62	762	75.43	112
	Comprehension	5689	711.09	713	50.59	51
	Oral	5689	725.83	731	60.61	65
	Total Writing	5689	711.69	718	47.53	56
	Total	5689	713.25	720	47.36	52

12	Listening	4969	715.81	725	52.62	70
	Writing Conventions	4969	714.08	712	44.55	50
	Reading	4969	718.20	718	48.75	58
	Speaking	4969	747.17	762	72.74	112
	Comprehension	4969	715.18	713	46.36	46
	Oral	4969	728.06	731	55.56	60
	Total Writing	4969	713.71	718	43.70	49
	Total	4969	716.01	720	41.75	50

Table 9.4: Descriptive Statistics for the Total Score without Reclassified Students by Grade

Grade	# of Items	Max Point	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Prop.	Proficient Cond. SEM
K	36	72	40363	49.84	14.78	524.81	52.44	0.80	6.68	0.69	18.27
1	76	100	18770	73.24	12.24	608.26	30.61	0.87	4.40	0.73	10.89
2	76	100	16628	83.93	11.42	643.06	40.89	0.90	3.53	0.84	11.97
3	76	100	15312	71.12	12.18	641.74	29.13	0.85	4.66	0.71	10.12
4	76	100	11536	76.28	11.98	655.77	31.99	0.87	4.28	0.76	10.50
5	76	100	9115	79.09	12.30	664.87	35.21	0.88	4.34	0.79	10.75
6	84	108	6435	84.44	15.17	688.91	38.58	0.92	4.34	0.78	8.75
7	84	108	6270	86.47	15.75	695.69	41.83	0.93	4.27	0.80	9.17
8	84	108	5573	88.84	15.47	703.46	43.28	0.95	3.58	0.82	9.70
9	84	108	6183	82.21	18.03	698.89	39.53	0.94	4.58	0.76	9.21
10	84	108	4599	83.28	16.67	700.47	37.17	0.91	4.97	0.77	9.38
11	84	108	3769	85.04	16.11	704.68	37.05	0.93	4.36	0.79	10.01
12	84	108	3558	87.09	14.33	708.92	34.47	0.91	4.25	0.81	10.01

Table 9.5: Descriptive Statistics for the Total Score without Reclassified Students by Grade and Time of Year

Grade	Term	N	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Prop.	Proficient Condl. SEM
K	Fall	18483	39.18	12.61	488.47	36.49	0.77	5.99	0.54	18.27
	Mid-Year	1251	51.06	13.66	528.29	50.44	0.83	5.59	0.71	
	Spring	20629	59.32	9.04	557.15	42.48	0.78	4.20	0.82	
1	Fall	1319	65.49	14.25	592.19	35.07	0.91	4.28	0.65	10.89
	Mid-Year	965	72.66	12.40	606.91	31.01	0.90	3.94	0.73	
	Spring	16486	73.90	11.83	609.62	29.82	0.86	4.42	0.74	
2	Fall	1425	77.21	14.12	622.80	42.64	0.92	3.88	0.77	11.97
	Mid-Year	828	84.18	11.53	643.82	40.47	0.91	3.37	0.84	
	Spring	14375	84.59	10.89	645.02	40.19	0.90	3.49	0.85	
3	Fall	1332	64.66	14.65	627.64	33.57	0.81	6.40	0.65	10.12
	Mid-Year	894	71.17	12.61	641.98	29.98	0.91	3.82	0.71	
	Spring	13086	71.78	11.68	643.16	28.20	0.85	4.49	0.72	
4	Fall	1216	71.72	14.04	644.48	34.72	0.87	4.97	0.72	10.50
	Mid-Year	702	76.87	12.02	657.85	33.42	0.85	4.71	0.77	
	Spring	9618	76.81	11.57	657.05	31.23	0.87	4.14	0.77	
5	Fall	1179	76.46	14.54	659.06	40.53	0.86	5.47	0.76	10.75
	Mid-Year	732	80.89	11.59	670.86	36.39	0.80	5.23	0.81	
	Spring	7204	79.34	11.91	665.21	34.00	0.89	4.00	0.79	
6	Fall	1005	83.24	16.77	687.38	42.31	0.90	5.28	0.77	8.75
	Mid-Year	760	87.23	13.83	696.32	38.48	0.93	3.65	0.81	
	Spring	4670	84.24	14.96	688.03	37.62	0.92	4.20	0.78	
7	Fall	1176	85.90	17.69	696.84	47.52	0.94	4.43	0.80	9.17
	Mid-Year	848	90.04	13.48	705.38	40.30	0.93	3.48	0.83	
	Spring	4246	85.92	15.50	693.44	40.14	0.92	4.36	0.80	
8	Fall	1025	87.70	16.62	701.27	46.36	0.95	3.66	0.81	9.7
	Mid-Year	750	91.51	13.84	711.28	42.55	0.94	3.36	0.85	
	Spring	3798	88.63	15.40	702.50	42.40	0.95	3.60	0.82	
9	Fall	2061	83.97	19.22	705.54	44.71	0.93	5.09	0.78	9.21
	Mid-Year	911	83.50	16.16	700.34	36.25	0.94	3.96	0.77	
	Spring	3211	80.71	17.61	694.21	36.10	0.94	4.38	0.75	
10	Fall	1158	82.56	18.46	700.18	40.64	0.95	4.05	0.76	9.38
	Mid-Year	706	85.88	14.58	705.79	34.36	0.89	4.88	0.80	
	Spring	2735	82.92	16.32	699.22	36.22	0.89	5.30	0.77	
11	Fall	987	84.92	17.11	705.46	39.48	0.95	3.92	0.79	10.01
	Mid-Year	717	86.99	14.87	708.82	35.35	0.94	3.75	0.81	
	Spring	2065	84.42	15.98	702.87	36.30	0.91	4.74	0.78	
12	Fall	1185	86.98	15.96	710.33	38.11	0.90	5.11	0.81	10.01
	Mid-Year	656	88.81	13.00	712.52	32.52	0.92	3.61	0.82	
	Spring	1717	86.51	13.56	706.56	32.34	0.92	3.80	0.80	

REFERENCES

- Andrew, B. J., & Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement* 36, 45-50.
- Andrich, A., & Luo, G. (2004). *Modern measurement and analysis in social science*. Perth, Western Australia: Murdoch University.
- Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service. (Reprint of chapter in R. L. Thorndike (Ed.). (1971). *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Brennan, R. L., & Lockwood, R. E. (1980). A comparison of the Nedelsky and Angoff cutting score procedures using generalizability theory. *Applied Psychological Measurement*, 4, 219-240.
- Cizek, G. J. (1996). Setting passing scores. *Educational Measurement: Issues and Practice*, 15(2), 20-31.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cross, L. H., Impara, J. C., Frary, R. B., & Jaeger, R. M. (1984). A comparison of three methods for establishing minimum standards on the National Teacher Examination. *Journal of Educational Measurement*, 21, 113-129.
- Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. New Jersey : Lawrence Erlbaum Associates, Publishers.
- Fitzpatrick, A. R. (1989). Social influences in standard-setting : The effects of social interaction on group judgments. *Review of Educational Research*, 59, 315-328.
- Haertel, E. H. (1996). *Estimating the decision consistency from a single administration of a performance assessment battery* (A report on the National Board of Professional Teaching Standards McGEN Assessment). Palo Alto, CA: Stanford University.
- Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8, 41-56.
- Hanson, B. A. (1995). USmooth: A program for smoothing univariate test score distribution (Version 1.5) [Computer software]. Iowa City, IA: American College Testing.

- Hanson, B. A. (1991). Method of moments estimates for the four-parameter beta compound binomial model and the calculation of classification consistency indexes (ACT Research Report 91-5). Iowa City, IA: American College Testing.
- Kolen, M. J. & Brennan R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.
- Linacre, J. M., & Wright, B. D. (2000). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago, IL: MESA Press
- Livingston, S. A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–198.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Nitko, A. J. (2004). *Educational Assessment of Students* (4th Ed.), Upper Saddle River, NJ: Person Education Inc.
- Office of English Language Acquisition Services (April, 2010). Test Accommodations for the Administration of the Arizona English Language Learner Assessment (AZELLA). Retrieved from <https://www.azed.gov/oelas/AZELLA/TestAccommodationsfortheAdministrationoftheAZELLA-April2010.pdf>
- Poggio, J. P., Glasnapp, D. R., & Eros, D. S. (1981). *An empirical investigation of the Angoff, Ebel, and Nedelsky standard setting methods*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (2000). *A survey and evaluation of recently developed procedures for setting standards on educational tests*. Paper written on contract to the National Assessment Governing Board.
- Skakun, E. N., & Kling, S. (1980). Comparability of methods for setting standards. *Journal of Educational Measurement*, 17, 229-235.
- Thissen, D. & Steinberg, L. (1983). *A response model for multiple choice items* (Psychometric Tech. Rep. No.1). Chicago, IL: National Opinion Research Center.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>

Young, M. J. & Yoon, B. (1998). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment*. (CSE Tech. Rep. 475). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation, Standards, and Student Testing.

**Arizona English Language Learner Assessment
(AZELLA) Form AZ-2 Technical Report**

Appendices A–C

2009 – 2010 Administration

February 2011

Prepared by

**Pearson
San Antonio, TX**

Table of Contents

APPENDIX A: ITEM-LEVEL STATISTICS BY LEVEL AND SUBTEST	2
A.1: Preliteracy	2
Listening	2
Prereading	2
Speaking.....	3
Pre-writing	3
A.2: Primary.....	4
Listening	4
Writing Conventions.....	4
Reading	5
Writing	5
Speaking.....	5
A.3: Elementary	6
Listening	6
Writing Conventions.....	6
Reading	7
Writing	7
Speaking.....	7
A.4: Middle Grades.....	8
Listening	8
Writing Conventions.....	8
Reading	9
Writing	9
Speaking.....	10
A.5: High School	11
Listening	11
Reading	12
Writing	12
Speaking.....	13
APPENDIX B: RAW SCORE-TO-SCALED SCORE CONVERSION TABLES BY STRAND.....	15
B.1: Preliteracy	15
B.2: Primary.....	22
B.3: Elementary	31
B.4: Middle Grades.....	40
B.5: High School.....	50
APPENDIX C: STANDARD SETTING MATERIALS	60
C.1: Standard Setting Meeting Two-Day Agenda	60
C.2: Performance Level Descriptors.....	61
Preliteracy (ELL I -- Kindergarten)	61
Primary (ELL II -- Grades 1–2)	65
Elementary (ELL III -- Grades 3–5)	70
Middle Grades (ELL IV -- Grades 6–8)	75
High School (ELL IV -- Grades 9–12)	79
C.3: AZELLA Form AZ-1 Standard Setting Summary Results in Raw Score Units.....	83

APPENDIX A: ITEM-LEVEL STATISTICS BY LEVEL AND SUBTEST

A.1: Preliteracy

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
1	MC	1	735	0.96	0.24	-5.94	0.19	0.98	0.75
2	MC	1	735	0.97	0.25	-6.28	0.23	0.95	0.57
3	MC	1	735	0.95	0.28	-5.73	0.18	0.96	0.69
4	MC	1	735	0.98	0.15	-6.71	0.27	1.01	0.82
5	MC	1	735	0.90	0.40	-5.15	0.14	1.18	0.90
6	MC	1	735	0.84	0.36	-4.25	0.11	0.98	0.87
7	MC	1	735	0.74	0.38	-3.74	0.09	1.07	0.98
8	MC	1	735	0.62	0.47	-3.16	0.09	1.00	0.95
9	MC	1	735	0.46	0.09	-2.15	0.08	1.30	1.40
10	MC	1	735	0.84	0.16	-4.26	0.11	1.12	1.44
11	MC	1	735	0.53	0.25	-2.51	0.08	1.14	1.19
12	MC	1	735	0.55	0.23	-2.60	0.08	1.16	1.24

Prereading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
13	MC	1	735	0.72	0.23	-3.46	0.09	1.14	1.18
14	MC	1	735	0.56	0.15	-2.64	0.08	1.23	1.31
15	MC	1	735	0.70	0.27	-3.38	0.09	1.11	1.14
16	MC	1	735	0.51	0.33	-2.41	0.08	1.06	1.06
17	MC	1	735	0.71	0.14	-3.42	0.09	1.22	1.33
18	MC	1	735	0.57	0.48	-2.68	0.08	0.91	0.90
19	MC	1	735	0.78	0.07	-3.82	0.10	1.26	1.42
20	MC	1	735	0.69	0.17	-3.28	0.09	1.20	1.40
21	MC	1	735	0.76	0.41	-3.48	0.09	0.89	0.79
22	MC	1	735	0.69	0.50	-3.61	0.09	1.03	0.97
23	MC	1	735	0.67	0.51	-3.26	0.09	0.91	0.84
24	MC	1	735	0.71	0.43	-3.26	0.09	0.93	0.85

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
25	SS	2	735	1.93	0.14	-4.79	0.11	0.83	2.69
26	SS	2	735	1.83	0.24	-5.22	0.10	1.09	1.27
27	SS	2	735	1.71	0.44	-3.92	0.07	0.89	0.86
28	SS	2	735	1.46	0.62	-3.44	0.06	0.84	0.82
29	SS	2	735	1.41	0.64	-3.27	0.06	0.83	0.78
30	SS	2	735	1.45	0.63	-3.35	0.06	0.84	0.80
31	SS	2	735	1.34	0.60	-3.09	0.06	0.91	0.89
32	SS	2	735	1.40	0.64	-3.33	0.06	0.82	0.79
33	SS	2	735	1.33	0.66	-3.11	0.06	0.81	0.80
34	SS	2	735	1.31	0.70	-3.06	0.06	0.75	0.71
35	SS	2	735	1.30	0.66	-3.07	0.06	0.81	0.78
36	SS	4	735	1.58	0.58	-1.87	0.04	1.29	1.30

Pre-writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
37	OE	1	735	0.96	0.18	-6.01	0.20	1.00	1.26
38	OE	1	735	0.93	0.23	-5.38	0.15	1.01	0.89
39	OE	1	735	0.97	0.16	-6.33	0.23	1.01	0.97
40	OE	1	735	0.89	0.36	-4.78	0.12	0.94	0.78
41	OE	1	735	0.87	0.34	-4.58	0.12	0.97	0.94
42	OE	2	735	1.43	0.48	-3.44	0.06	1.04	1.00
43	OE	2	735	1.27	0.49	-3.02	0.06	1.04	1.01
44	OE	2	735	1.36	0.48	-3.26	0.06	1.04	1.02
45	OE	2	735	1.29	0.54	-3.04	0.06	0.98	0.94
46	OE	1	735	0.86	0.35	-4.49	0.11	1.01	0.98
47	OE	1	735	0.45	0.33	-1.80	0.08	1.11	1.14
48	OE	1	735	0.92	0.33	-5.69	0.17	1.48	1.02
49	OE	1	735	0.63	0.48	-3.16	0.09	0.97	0.89
50	OE	1	735	0.50	0.39	-2.35	0.08	1.01	1.02
51	OE	1	735	0.41	0.52	-1.89	0.08	0.85	0.81
52	OE	1	735	0.36	0.44	-1.67	0.08	0.93	0.92
53	OE	2	735	0.10	0.28	0.66	0.12	0.99	0.89

A.2: Primary

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
1	MC	1	639	0.90	0.24	-2.22	0.14	1.01	1.06
2	MC	1	639	0.74	0.33	-0.98	0.10	1.04	1.15
3	MC	1	639	0.85	0.36	-1.75	0.12	0.97	0.79
4	MC	1	639	0.77	0.32	-1.14	0.10	1.04	1.01
5	MC	1	639	0.97	0.24	-3.62	0.24	0.94	0.79
6	MC	1	639	0.97	0.26	-3.56	0.23	0.94	0.61
7	MC	1	639	0.98	0.20	-3.87	0.26	0.96	0.57
8	MC	1	639	0.98	0.21	-4.26	0.32	1.01	0.50
9	MC	1	639	0.97	0.17	-3.38	0.21	0.76	0.88
10	MC	1	639	0.99	0.20	-4.10	0.29	0.72	0.32
11	MC	1	639	0.99	0.16	-4.18	0.30	0.51	0.28
12	MC	1	639	0.97	0.17	-3.51	0.23	0.99	1.10
13	MC	1	639	0.89	0.23	-2.15	0.13	1.04	1.16
14	MC	1	639	0.96	0.20	-3.20	0.20	0.98	0.81
15	MC	1	639	0.89	0.23	-2.10	0.13	1.02	1.25
16	MC	1	639	0.84	0.22	-1.64	0.11	1.08	1.23
17	MC	1	639	0.89	0.24	-2.15	0.13	1.02	1.21
18	MC	1	639	0.79	0.24	-1.32	0.10	1.09	1.35
19	MC	1	639	0.89	0.24	-1.97	0.12	0.94	1.09
20	MC	1	639	0.86	0.24	-1.69	0.11	0.96	1.03

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
21	MC	1	639	0.79	0.43	-1.26	0.10	0.95	0.74
22	MC	1	639	0.72	0.50	-0.85	0.10	0.90	0.78
23	MC	1	639	0.68	0.62	-0.63	0.09	0.78	0.67
24	MC	1	639	0.87	0.43	-1.91	0.12	0.91	0.65
25	MC	1	639	0.93	0.34	-2.72	0.16	0.92	0.68
26	MC	1	639	0.55	0.38	0.05	0.09	1.04	1.06
27	MC	1	639	0.76	0.45	-1.11	0.10	0.93	0.77
28	MC	1	639	0.48	0.26	0.42	0.09	1.21	1.36
29	MC	1	639	0.75	0.49	-1.01	0.10	0.90	0.76
30	MC	1	639	0.58	0.45	-0.12	0.09	0.97	0.92
31	MC	1	639	0.54	0.46	0.10	0.09	0.96	0.93
32	MC	1	639	0.61	0.45	-0.45	0.09	0.99	0.96
33	MC	1	639	0.82	0.43	-0.86	0.10	0.75	0.61
34	MC	1	639	0.59	0.37	-0.16	0.09	1.06	1.07
35	MC	1	639	0.58	0.41	-0.08	0.09	1.02	1.01
36	MC	1	639	0.78	0.45	-0.47	0.09	0.81	0.72
37	MC	1	639	0.63	0.45	-0.36	0.09	0.96	0.91
38	MC	1	639	0.59	0.46	-0.14	0.09	0.95	0.91
39	MC	1	639	0.66	0.31	-0.51	0.09	1.10	1.19
40	MC	1	639	0.71	0.54	-0.60	0.09	0.82	0.72

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch		
							Diff	INFIT	OUTFIT
41	MC	1	639	0.90	0.40	-2.26	0.14	0.91	0.56
42	MC	1	639	0.89	0.45	-2.13	0.13	0.89	0.54
43	MC	1	639	0.89	0.38	-2.19	0.13	0.93	0.65
44	MC	1	639	0.76	0.40	-1.07	0.10	0.98	0.93
45	MC	1	639	0.72	0.45	-0.85	0.10	0.94	0.88
46	MC	1	639	0.79	0.54	-1.32	0.10	0.84	0.62
47	MC	1	639	0.66	0.07	-0.50	0.09	1.34	1.58
48	MC	1	639	0.68	0.38	-1.08	0.10	1.22	1.19
49	MC	1	639	0.75	0.50	-0.89	0.10	0.84	0.68
50	MC	1	639	0.67	0.63	-0.40	0.09	0.74	0.64
51	MC	1	639	0.60	0.32	-0.56	0.09	1.18	1.17
52	MC	1	639	0.56	0.46	0.01	0.09	0.96	0.93
53	MC	1	639	0.56	0.43	-0.01	0.09	1.00	1.00
54	MC	1	639	0.46	0.24	0.14	0.09	1.23	1.33
55	MC	1	639	0.54	0.36	0.08	0.09	1.07	1.07
56	MC	1	639	0.50	0.40	0.32	0.09	0.99	1.03
57	MC	1	639	0.64	0.37	-0.42	0.09	1.04	1.05
58	MC	1	639	0.66	0.51	-0.52	0.09	0.90	0.82
59	MC	1	639	0.38	0.40	0.14	0.09	1.07	1.10
60	MC	1	639	0.46	0.43	0.52	0.09	0.99	1.03

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch		
							Diff	INFIT	OUTFIT
61	OE	4	639	2.21	0.70	-0.18	0.06	0.81	0.80
62	OE	4	639	1.97	0.72	0.54	0.06	0.77	0.77

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch		
							Diff	INFIT	OUTFIT
63	SS	2	639	1.75	0.41	-1.62	0.08	0.98	1.20
64	SS	2	639	1.66	0.45	-1.47	0.08	0.98	1.09
65	SS	2	639	1.74	0.45	-2.13	0.09	0.93	0.96
66	SS	2	639	1.56	0.49	-1.32	0.07	0.97	1.07
67	SS	2	639	1.09	0.38	0.08	0.06	1.29	1.30
68	SS	2	639	1.72	0.46	-1.47	0.08	0.81	0.81
69	SS	2	639	1.61	0.46	-1.56	0.08	0.96	1.30
70	SS	2	639	1.46	0.34	-1.09	0.07	1.31	1.82
71	SS	2	639	1.65	0.49	-1.40	0.07	0.85	0.95
72	SS	2	639	1.54	0.49	-1.60	0.08	0.97	0.99
73	SS	2	639	1.61	0.46	-1.28	0.07	1.01	1.16
74	SS	2	639	1.74	0.41	-2.22	0.09	0.96	1.01
75	SS	4	639	2.62	0.59	-0.37	0.04	1.24	1.23
76	SS	4	639	2.69	0.61	-0.61	0.05	1.16	1.23

A.3: Elementary

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
1	MC	1	1004	0.98	0.09	-2.75	0.21	0.69	0.90
2	MC	1	1004	0.97	0.26	-2.18	0.16	0.68	0.49
3	MC	1	1004	0.97	0.10	-2.47	0.18	1.02	1.15
4	MC	1	1004	0.97	0.22	-2.47	0.18	0.95	0.92
5	MC	1	1004	0.77	0.26	-0.09	0.08	1.04	1.14
6	MC	1	1004	0.29	0.08	2.31	0.07	1.18	1.38
7	MC	1	1004	0.74	0.27	0.13	0.08	1.05	1.12
8	MC	1	1004	0.76	0.41	0.00	0.08	0.93	0.93
9	MC	1	1004	0.85	0.31	-0.37	0.09	0.84	0.87
10	MC	1	1004	0.65	0.36	0.45	0.07	1.04	1.10
11	MC	1	1004	0.71	0.34	0.29	0.07	1.00	0.97
12	MC	1	1004	0.64	0.33	0.71	0.07	1.01	1.00
13	MC	1	1004	0.63	0.41	0.71	0.07	0.95	0.92
14	MC	1	1004	0.64	0.25	0.63	0.07	1.08	1.14
15	MC	1	1004	0.64	0.36	0.65	0.07	0.99	0.97
16	MC	1	1004	0.97	0.16	-2.54	0.19	0.99	0.90
17	MC	1	1004	0.85	0.30	-0.65	0.09	0.99	1.00
18	MC	1	1004	0.30	0.20	2.27	0.07	1.09	1.21
19	MC	1	1004	0.64	0.29	0.66	0.07	1.05	1.08
20	MC	1	1004	0.61	0.32	0.80	0.07	1.03	1.02

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
21	MC	1	1004	0.95	0.26	-2.00	0.15	0.97	0.71
22	MC	1	1004	0.91	0.31	-1.27	0.11	0.96	0.83
23	MC	1	1004	0.98	0.25	-2.70	0.20	0.95	0.54
24	MC	1	1004	0.92	0.18	-1.00	0.10	0.77	0.97
25	MC	1	1004	0.88	0.37	-0.89	0.10	0.93	0.80
26	MC	1	1004	0.66	0.45	0.56	0.07	0.92	0.87
27	MC	1	1004	0.70	0.40	0.36	0.07	0.94	0.88
28	MC	1	1004	0.64	0.42	0.61	0.07	0.95	0.94
29	MC	1	1004	0.63	0.44	0.63	0.07	0.94	0.90
30	MC	1	1004	0.64	0.48	0.65	0.07	0.90	0.86
31	MC	1	1004	0.59	0.38	0.91	0.07	0.98	0.98
32	MC	1	1004	0.55	0.32	1.07	0.07	1.02	1.05
33	MC	1	1004	0.51	0.34	1.28	0.07	1.01	0.99
34	MC	1	1004	0.92	0.25	-1.38	0.12	1.00	0.90
35	MC	1	1004	0.87	0.42	-0.81	0.10	0.89	0.77
36	MC	1	1004	0.80	0.48	-0.30	0.08	0.87	0.75
37	MC	1	1004	0.60	0.34	0.84	0.07	1.01	1.01
38	MC	1	1004	0.68	0.38	0.43	0.07	0.97	0.98
39	MC	1	1004	0.63	0.29	0.69	0.07	1.04	1.05
40	MC	1	1004	0.58	0.41	0.92	0.07	0.96	0.95

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
41	MC	1	1004	0.96	0.31	-2.20	0.16	0.92	0.71
42	MC	1	1004	0.96	0.26	-2.17	0.16	0.96	0.72
43	MC	1	1004	0.92	0.36	-1.80	0.14	1.30	1.13
44	MC	1	1004	0.47	0.37	1.20	0.07	0.99	1.01
45	MC	1	1004	0.41	0.26	1.84	0.07	1.10	1.16
46	MC	1	1004	0.41	0.34	1.48	0.07	0.99	0.99
47	MC	1	1004	0.44	0.33	1.57	0.07	1.01	1.02
48	MC	1	1004	0.52	0.29	1.23	0.07	1.05	1.06
49	MC	1	1004	0.54	0.43	1.11	0.07	0.94	0.92
50	MC	1	1004	0.62	0.39	0.76	0.07	0.97	0.96
51	MC	1	1004	0.54	0.35	1.13	0.07	1.00	1.00
52	MC	1	1004	0.41	0.34	1.71	0.07	0.99	1.02
53	MC	1	1004	0.40	0.38	1.75	0.07	0.95	0.98
54	MC	1	1004	0.25	0.16	2.65	0.08	1.12	1.35
55	MC	1	1004	0.38	0.30	1.71	0.07	1.00	1.04
56	MC	1	1004	0.62	0.35	0.73	0.07	1.00	0.99
57	MC	1	1004	0.62	0.38	0.76	0.07	0.97	0.95
58	MC	1	1004	0.50	0.41	1.28	0.07	0.95	0.93
59	MC	1	1004	0.49	0.25	1.34	0.07	1.09	1.12
60	MC	1	1004	0.42	0.19	1.66	0.07	1.11	1.21

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
61	OE	4	1004	2.44	0.59	0.37	0.04	0.71	0.75
62	OE	4	1004	2.34	0.59	1.07	0.06	0.83	0.84

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
63	SS	2	1004	1.81	0.32	-0.85	0.07	1.02	1.13
64	SS	2	1004	1.65	0.44	-0.41	0.06	0.96	0.98
65	SS	2	1004	1.77	0.42	-1.54	0.08	0.93	0.88
66	SS	2	1004	1.78	0.39	-1.63	0.08	0.95	0.90
67	SS	2	1004	1.77	0.39	-1.28	0.08	0.95	0.93
68	SS	4	1004	3.45	0.47	-1.11	0.05	0.97	0.99
69	SS	4	1004	3.33	0.49	-0.21	0.04	1.06	1.30
70	SS	2	1004	1.73	0.42	-0.83	0.07	0.95	0.91
71	SS	2	1004	1.60	0.42	-0.36	0.06	0.99	1.00
72	SS	2	1004	1.63	0.44	-0.47	0.06	0.96	0.97
73	SS	2	1004	1.72	0.39	-1.09	0.07	0.97	0.99
74	SS	2	1004	1.82	0.33	-0.83	0.07	0.78	0.75
75	SS	2	1004	1.69	0.41	-0.26	0.06	0.78	0.83
76	SS	2	1004	1.76	0.40	-1.18	0.07	0.95	0.97

A.4: Middle Grades

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
1	MC	1	1076	0.99	0.11	-2.22	0.24	0.64	0.46
2	MC	1	1076	0.94	0.30	-1.57	0.18	1.82	1.82
3	MC	1	1076	0.93	0.36	-0.73	0.13	0.95	0.68
4	MC	1	1076	0.83	0.37	0.47	0.09	0.98	0.94
5	MC	1	1076	0.87	0.43	0.09	0.10	0.92	0.76
6	MC	1	1076	0.77	0.32	0.87	0.08	1.05	1.07
7	MC	1	1076	0.73	0.36	1.12	0.07	1.02	1.01
8	MC	1	1076	0.75	0.16	0.48	0.09	1.53	2.08
9	MC	1	1076	0.53	0.36	2.17	0.07	1.02	1.05
10	MC	1	1076	0.87	0.36	0.10	0.10	0.97	0.96
11	MC	1	1076	0.73	0.37	1.14	0.07	1.01	1.03
12	MC	1	1076	0.87	0.37	0.07	0.10	0.97	0.86
13	MC	1	1076	0.52	0.29	2.21	0.07	1.08	1.14
14	MC	1	1076	0.79	0.29	0.75	0.08	1.06	1.28
15	MC	1	1076	0.51	0.20	2.28	0.07	1.19	1.28
16	MC	1	1076	0.75	0.38	0.99	0.08	1.00	0.99
17	MC	1	1076	0.65	0.36	1.59	0.07	1.03	1.02
18	MC	1	1076	0.60	0.26	1.85	0.07	1.13	1.22
19	MC	1	1076	0.85	0.31	0.05	0.10	1.16	1.24
20	MC	1	1076	0.68	0.25	1.39	0.07	1.14	1.22

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
21	MC	1	1076	0.90	0.39	-0.42	0.11	1.04	0.85
22	MC	1	1076	0.96	0.29	-1.45	0.17	1.25	0.97
23	MC	1	1076	0.91	0.31	-0.25	0.11	0.96	0.92
24	MC	1	1076	0.94	0.22	-0.86	0.13	1.01	1.32
25	MC	1	1076	0.93	0.22	-0.60	0.12	1.04	1.20
26	MC	1	1076	0.87	0.30	0.10	0.10	1.03	0.99
27	MC	1	1076	0.83	0.20	0.46	0.09	1.13	1.23
28	MC	1	1076	0.85	0.39	0.32	0.09	0.92	0.84
29	MC	1	1076	0.85	0.43	0.05	0.10	1.12	1.00
30	MC	1	1076	0.83	0.45	0.46	0.09	0.91	0.90
31	MC	1	1076	0.76	0.43	0.95	0.08	0.96	0.88
32	MC	1	1076	0.76	0.34	0.97	0.08	1.03	1.08
33	MC	1	1076	0.69	0.42	1.36	0.07	0.96	0.91
34	MC	1	1076	0.65	0.37	1.58	0.07	1.01	1.00
35	MC	1	1076	0.62	0.36	1.75	0.07	1.03	1.03
36	MC	1	1076	0.51	0.35	2.26	0.07	1.03	1.05
37	MC	1	1076	0.68	0.44	1.73	0.07	0.89	0.85
38	MC	1	1076	0.78	0.30	0.40	0.09	1.34	1.43
39	MC	1	1076	0.74	0.42	1.05	0.08	0.97	0.91
40	MC	1	1076	0.85	0.48	0.24	0.09	0.90	0.71

41	MC	1	1076	0.89	0.45	-0.07	0.10	0.90	0.70
42	MC	1	1076	0.68	0.29	1.39	0.07	1.09	1.10
43	MC	1	1076	0.61	0.32	1.76	0.07	1.07	1.09
44	MC	1	1076	0.70	0.40	1.32	0.07	0.99	0.99

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
45	MC	1	1076	0.86	0.36	0.56	0.08	0.83	0.88
46	MC	1	1076	0.96	0.38	-0.82	0.13	0.55	0.29
47	MC	1	1076	0.88	0.36	0.27	0.09	0.81	0.74
48	MC	1	1076	0.90	0.39	-0.25	0.11	0.94	0.78
49	MC	1	1076	0.70	0.50	1.30	0.07	0.90	0.86
50	MC	1	1076	0.75	0.40	1.02	0.08	0.98	0.95
51	MC	1	1076	0.67	0.47	1.49	0.07	0.92	0.86
52	MC	1	1076	0.88	0.37	-0.04	0.10	0.96	0.87
53	MC	1	1076	0.72	0.51	1.21	0.07	0.89	0.83
54	MC	1	1076	0.68	0.43	1.42	0.07	0.97	0.95
55	MC	1	1076	0.65	0.35	1.60	0.07	1.04	1.04
56	MC	1	1076	0.58	0.32	1.48	0.07	1.20	1.24
57	MC	1	1076	0.55	0.25	2.06	0.07	1.12	1.14
58	MC	1	1076	0.35	0.16	3.05	0.07	1.15	1.34
59	MC	1	1076	0.57	0.39	1.31	0.07	1.27	1.30
60	MC	1	1076	0.85	0.39	0.27	0.09	0.97	0.83
61	MC	1	1076	0.83	0.36	0.49	0.09	1.01	0.89
62	MC	1	1076	0.62	0.39	1.75	0.07	1.00	1.01
63	MC	1	1076	0.81	0.35	0.62	0.08	1.02	0.91
64	MC	1	1076	0.72	0.45	1.16	0.07	0.94	0.82
65	MC	1	1076	0.62	0.36	1.71	0.07	1.03	1.02
66	MC	1	1076	0.61	0.48	1.77	0.07	0.91	0.88
67	MC	1	1076	0.67	0.26	1.47	0.07	1.13	1.21
68	MC	1	1076	0.55	0.26	2.09	0.07	1.12	1.15

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
69	OE	4	1076	2.52	0.46	1.63	0.05	1.02	1.07
70	OE	4	1076	2.45	0.47	1.77	0.05	1.01	1.04

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Bisequal	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
71	SS	2	1076	1.66	0.48	0.35	0.06	0.97	0.98
72	SS	2	1076	1.72	0.47	-0.12	0.07	0.95	0.90
73	SS	2	1076	1.64	0.51	0.68	0.06	0.98	0.94
74	SS	2	1076	1.66	0.55	0.59	0.06	0.91	0.85
75	SS	2	1076	1.65	0.54	0.53	0.06	0.92	0.86
76	SS	2	1076	1.51	0.57	0.85	0.06	0.89	0.85
77	SS	2	1076	1.57	0.60	0.74	0.06	0.85	0.82
78	SS	2	1076	1.57	0.59	0.66	0.06	0.86	0.83
79	SS	4	1076	3.11	0.60	1.02	0.04	1.06	1.06
80	SS	2	1076	1.65	0.51	0.60	0.06	0.78	0.75
81	SS	2	1076	1.80	0.46	0.25	0.06	0.76	0.68
82	SS	2	1076	1.69	0.52	0.61	0.06	0.76	0.71
83	SS	2	1076	1.75	0.49	0.37	0.06	0.76	0.78
84	SS	4	1076	3.26	0.60	0.87	0.04	0.78	0.76

A.5: High School

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
1	MC	1	1489	0.81	0.43	0.86	0.07	0.95	0.86
2	MC	1	1489	0.95	0.42	-0.77	0.12	0.87	0.56
3	MC	1	1489	0.94	0.38	-0.70	0.12	0.91	0.69
4	MC	1	1489	0.85	0.33	0.48	0.08	1.01	1.07
5	MC	1	1489	0.91	0.34	-0.15	0.10	0.96	0.95
6	MC	1	1489	0.84	0.32	0.57	0.08	1.03	1.05
7	MC	1	1489	0.54	0.32	2.34	0.06	1.06	1.10
8	MC	1	1489	0.79	0.32	0.97	0.07	1.05	1.05
9	MC	1	1489	0.66	0.43	1.78	0.06	0.97	0.94
10	MC	1	1489	0.84	0.56	0.55	0.08	0.82	0.64
11	MC	1	1489	0.67	0.29	1.69	0.06	1.10	1.14
12	MC	1	1489	0.77	0.53	1.14	0.07	0.87	0.81
13	MC	1	1489	0.65	0.48	1.81	0.06	0.92	0.89
14	MC	1	1489	0.47	0.21	2.70	0.06	1.15	1.26
15	MC	1	1489	0.71	0.45	2.06	0.06	0.87	0.84
16	MC	1	1489	0.80	0.23	0.93	0.07	1.13	1.26
17	MC	1	1489	0.85	0.38	0.50	0.08	0.96	1.00
18	MC	1	1489	0.82	0.42	0.79	0.07	0.95	0.92
19	MC	1	1489	0.52	0.34	2.46	0.06	1.03	1.06
20	MC	1	1489	0.83	0.23	0.65	0.07	1.11	1.24

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
21	MC	1	1489	0.94	0.39	-0.59	0.11	0.91	0.61
22	MC	1	1489	0.91	0.36	-0.13	0.09	0.96	0.83
23	MC	1	1489	0.97	0.14	-1.27	0.15	1.02	1.60
24	MC	1	1489	0.92	0.20	-0.28	0.10	1.05	1.21
25	MC	1	1489	0.82	0.22	0.72	0.07	1.12	1.25
26	MC	1	1489	0.69	0.13	1.59	0.06	1.25	1.31
27	MC	1	1489	0.69	0.34	1.59	0.06	1.05	1.03
28	MC	1	1489	0.58	0.18	2.18	0.06	1.21	1.28
29	MC	1	1489	0.90	0.29	0.01	0.09	1.01	1.01
30	MC	1	1489	0.95	0.43	-0.77	0.12	0.86	0.48
31	MC	1	1489	0.84	0.51	1.02	0.07	0.70	0.59
32	MC	1	1489	0.72	0.31	1.41	0.06	1.07	1.05
33	MC	1	1489	0.80	0.33	1.70	0.06	0.89	0.84
34	MC	1	1489	0.67	0.40	1.73	0.06	1.00	0.94
35	MC	1	1489	0.53	0.27	2.38	0.06	1.11	1.15
36	MC	1	1489	0.52	0.34	2.47	0.06	1.03	1.08
37	MC	1	1489	0.91	0.37	-0.17	0.10	0.94	0.82
38	MC	1	1489	0.76	0.57	1.18	0.07	0.83	0.72
39	MC	1	1489	0.66	0.47	1.77	0.06	0.93	0.93
40	MC	1	1489	0.56	0.45	2.28	0.06	0.95	0.95

41	MC	1	1489	0.56	0.11	2.28	0.06	1.27	1.37
42	MC	1	1489	0.46	0.25	2.76	0.06	1.12	1.22
43	MC	1	1489	0.58	0.52	2.55	0.06	0.88	0.86
44	MC	1	1489	0.61	0.41	2.06	0.06	0.97	0.98

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
45	MC	1	1489	0.99	0.09	-2.40	0.25	1.01	1.05
46	MC	1	1489	0.89	0.39	-0.02	0.09	1.03	0.90
47	MC	1	1489	0.92	0.39	0.06	0.09	0.75	0.64
48	MC	1	1489	0.73	0.40	1.37	0.06	0.99	0.98
49	MC	1	1489	0.69	0.39	1.59	0.06	1.00	0.97
50	MC	1	1489	0.61	0.37	2.01	0.06	1.03	1.01
51	MC	1	1489	0.76	0.26	1.20	0.07	1.11	1.21
52	MC	1	1489	0.76	0.53	1.17	0.07	0.87	0.75
53	MC	1	1489	0.76	0.25	1.16	0.07	1.12	1.21
54	MC	1	1489	0.58	0.41	2.14	0.06	0.98	0.98
55	MC	1	1489	0.82	0.29	0.77	0.07	1.07	1.11
56	MC	1	1489	0.78	0.28	1.03	0.07	1.09	1.13
57	MC	1	1489	0.77	0.45	1.13	0.07	0.94	0.87
58	MC	1	1489	0.75	0.41	1.22	0.07	0.98	0.94
59	MC	1	1489	0.76	0.48	1.15	0.07	0.91	0.80
60	MC	1	1489	0.39	0.08	3.06	0.06	1.22	1.48
61	MC	1	1489	0.48	0.11	2.63	0.06	1.26	1.37
62	MC	1	1489	0.45	0.32	2.79	0.06	1.05	1.12
63	MC	1	1489	0.49	0.36	2.58	0.06	1.00	1.04
64	MC	1	1489	0.41	0.20	2.37	0.06	1.22	1.29
65	MC	1	1489	0.57	0.46	2.21	0.06	0.93	0.90
66	MC	1	1489	0.59	0.30	1.47	0.06	1.37	1.45
67	MC	1	1489	0.64	0.32	1.87	0.06	1.07	1.07
68	MC	1	1489	0.60	0.38	2.05	0.06	1.01	0.99

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
69	OE	4	1489	2.87	0.57	1.28	0.03	0.72	0.74
70	OE	4	1489	2.71	0.53	1.77	0.04	1.12	1.23

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	INFIT	OUTFIT
71	SS	2	1489	1.73	0.54	0.29	0.06	0.88	0.84
72	SS	2	1489	1.60	0.63	1.02	0.05	0.82	0.75
73	SS	2	1489	1.53	0.66	1.31	0.04	0.80	0.77
74	SS	2	1489	1.56	0.60	1.10	0.05	0.88	0.83
75	SS	2	1489	1.38	0.63	1.77	0.04	0.89	0.84
76	SS	2	1489	2.84	0.71	1.69	0.03	0.93	0.93
77	SS	2	1489	1.60	0.66	1.03	0.05	0.78	0.75
78	SS	2	1489	1.22	0.61	1.94	0.04	0.86	0.85
79	SS	2	1489	1.54	0.64	1.19	0.05	0.83	0.80
80	SS	2	1489	2.59	0.72	1.96	0.03	0.92	0.93
81	SS	2	1489	1.65	0.63	0.69	0.05	0.77	0.74
82	SS	4	1489	1.51	0.63	1.04	0.05	0.95	0.88
83	SS	4	1489	1.54	0.61	0.75	0.05	0.93	0.99
84	SS	4	1489	1.72	0.59	0.62	0.05	0.84	0.78

APPENDIX B:

RAW SCORE-TO-SCALED SCORE CONVERSION TABLES BY STRAND

APPENDIX B: RAW SCORE-TO-SCALED SCORE CONVERSION TABLES BY STRAND

B.1: Preliteracy

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-9.14	2.05	300	71.82
1	-7.59	1.11	334	38.75
2	-6.66	0.87	367	30.28
3	-5.99	0.78	390	27.27
4	-5.42	0.74	410	25.94
5	-4.88	0.72	429	25.31
6	-4.36	0.72	447	25.10
7	-3.85	0.72	465	25.20
8	-3.32	0.74	484	25.76
9	-2.75	0.77	504	27.09
10	-2.09	0.86	527	30.14
11	-1.17	1.10	559	38.61
12	0.37	2.05	613	71.72

Strand 3: Prereading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-7.15	2.02	350	70.81
1	-5.69	1.05	401	36.72
2	-4.89	0.78	429	27.34
3	-4.36	0.68	447	23.63
4	-3.95	0.62	462	21.77
5	-3.58	0.60	475	20.86
6	-3.23	0.59	487	20.62
7	-2.88	0.60	499	20.90
8	-2.51	0.62	512	21.84
9	-2.09	0.68	527	23.73
10	-1.56	0.79	545	27.48
11	-0.75	1.05	574	36.82
12	0.71	2.03	625	70.88

Strand 4: Prewriting				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-8.97	2.04	300	71.37
1	-7.47	1.08	339	37.73
2	-6.61	0.82	369	28.60
3	-6.03	0.71	389	24.85
4	-5.57	0.65	405	22.72
5	-5.18	0.61	419	21.35
6	-4.82	0.58	431	20.37
7	-4.50	0.56	443	19.64
8	-4.19	0.55	453	19.08
9	-3.90	0.53	464	18.69
10	-3.62	0.53	473	18.45
11	-3.34	0.52	483	18.34
12	-3.07	0.53	493	18.38
13	-2.79	0.53	502	18.62
14	-2.50	0.54	513	19.04
15	-2.19	0.56	523	19.74
16	-1.86	0.59	535	20.79
17	-1.48	0.64	548	22.33
18	-1.03	0.70	564	24.57
19	-0.48	0.79	583	27.62
20	0.23	0.89	608	31.29
21	1.19	1.10	642	38.50
22	2.68	2.02	694	70.56

Strand 5: Speaking				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-8.23	2.03	312	71.12
1	-6.76	1.05	363	36.65
2	-5.99	0.76	390	26.43
3	-5.52	0.63	407	21.95
4	-5.17	0.56	419	19.46
5	-4.89	0.51	429	17.96
6	-4.64	0.49	438	16.98
7	-4.41	0.47	446	16.28
8	-4.20	0.45	453	15.75
9	-4.01	0.44	460	15.30
10	-3.82	0.43	466	14.95
11	-3.64	0.42	473	14.70
12	-3.46	0.41	479	14.49
13	-3.29	0.41	485	14.39
14	-3.12	0.41	491	14.35
15	-2.96	0.41	497	14.39
16	-2.79	0.41	503	14.49
17	-2.61	0.42	509	14.74
18	-2.43	0.43	515	15.05
19	-2.24	0.44	522	15.51
20	-2.03	0.46	529	16.21
21	-1.81	0.49	537	17.22
22	-1.54	0.54	546	18.80
23	-1.22	0.61	557	21.25
24	-0.78	0.73	573	25.62
25	-0.05	1.02	598	35.67
26	1.37	2.01	648	70.39

Strand 10: Comprehension (Listening + Pre-reading)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-9.28	2.04	300	71.40
1	-7.77	1.08	328	37.77
2	-6.91	0.82	358	28.53
3	-6.34	0.70	378	24.64
4	-5.89	0.64	394	22.37
5	-5.51	0.59	407	20.76
6	-5.18	0.56	419	19.53
7	-4.88	0.53	429	18.55
8	-4.61	0.51	438	17.78
9	-4.36	0.49	447	17.19
10	-4.13	0.48	455	16.73
11	-3.91	0.47	463	16.42
12	-3.69	0.46	471	16.17
13	-3.48	0.46	478	16.10
14	-3.26	0.46	486	16.10
15	-3.05	0.46	493	16.24
16	-2.83	0.47	501	16.52
17	-2.60	0.49	509	16.98
18	-2.36	0.50	518	17.64
19	-2.09	0.53	527	18.62
20	-1.78	0.57	538	20.09
21	-1.42	0.64	550	22.40
22	-0.94	0.76	567	26.50
23	-0.17	1.04	594	36.23
24	1.27	2.02	644	70.60

Strand 11: Oral (Listening + Speaking)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-9.50	2.03	300	70.98
1	-8.02	1.06	319	36.96
2	-7.22	0.78	347	27.37
3	-6.70	0.66	365	23.21
4	-6.31	0.59	379	20.72
5	-5.99	0.54	390	18.97
6	-5.72	0.50	400	17.64
7	-5.48	0.48	408	16.63
8	-5.26	0.45	416	15.82
9	-5.07	0.43	423	15.16
10	-4.88	0.42	429	14.63
11	-4.71	0.41	435	14.25
12	-4.55	0.40	441	13.90
13	-4.40	0.39	446	13.62
14	-4.25	0.38	451	13.41
15	-4.10	0.38	456	13.20
16	-3.96	0.37	461	13.02
17	-3.83	0.37	466	12.88
18	-3.69	0.36	471	12.74
19	-3.56	0.36	475	12.67
20	-3.43	0.36	480	12.60
21	-3.30	0.36	485	12.57
22	-3.17	0.36	489	12.57
23	-3.04	0.36	494	12.60
24	-2.91	0.36	498	12.67
25	-2.78	0.37	503	12.78
26	-2.64	0.37	507	12.95
27	-2.50	0.38	512	13.16
28	-2.36	0.38	517	13.44
29	-2.21	0.39	523	13.79
30	-2.05	0.41	528	14.28
31	-1.87	0.43	534	14.91
32	-1.68	0.45	541	15.75
33	-1.46	0.48	549	16.94
34	-1.21	0.53	558	18.62
35	-0.89	0.61	569	21.18
36	-0.45	0.73	584	25.59
37	0.28	1.02	610	35.63
38	1.70	2.01	659	70.32

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-10.04	2.02	300	70.53
1	-8.61	1.03	300	36.02
2	-7.85	0.75	325	26.15
3	-7.39	0.63	341	21.88
4	-7.04	0.55	353	19.39
5	-6.76	0.51	363	17.68
6	-6.53	0.47	372	16.45
7	-6.32	0.44	379	15.47
8	-6.13	0.42	385	14.70
9	-5.96	0.40	391	14.04
10	-5.81	0.39	397	13.48
11	-5.67	0.37	402	12.99
12	-5.53	0.36	406	12.57
13	-5.41	0.35	411	12.22
14	-5.29	0.34	415	11.87
15	-5.18	0.33	419	11.59
16	-5.07	0.32	423	11.31
17	-4.97	0.32	426	11.10
18	-4.87	0.31	430	10.89
19	-4.77	0.31	433	10.71
20	-4.68	0.30	436	10.54
21	-4.59	0.30	439	10.40
22	-4.50	0.29	442	10.26
23	-4.42	0.29	445	10.12
24	-4.33	0.29	448	10.01
25	-4.25	0.28	451	9.91
26	-4.17	0.28	454	9.80
27	-4.10	0.28	457	9.73
28	-4.02	0.28	459	9.66
29	-3.94	0.27	462	9.59
30	-3.87	0.27	465	9.52
31	-3.80	0.27	467	9.45
32	-3.72	0.27	470	9.42
33	-3.65	0.27	472	9.38
34	-3.58	0.27	475	9.31
35	-3.51	0.27	477	9.31
36	-3.44	0.27	480	9.28
37	-3.37	0.27	482	9.28
38	-3.30	0.26	485	9.24
39	-3.23	0.26	487	9.24
40	-3.16	0.26	490	9.24
41	-3.09	0.27	492	9.28
42	-3.02	0.27	494	9.28

43	-2.94	0.27	497	9.31
44	-2.87	0.27	499	9.35
45	-2.80	0.27	502	9.42
46	-2.73	0.27	505	9.45
47	-2.65	0.27	507	9.52
48	-2.58	0.27	510	9.59
49	-2.50	0.28	512	9.70
50	-2.43	0.28	515	9.80
51	-2.35	0.28	518	9.91
52	-2.26	0.29	521	10.05
53	-2.18	0.29	524	10.19
54	-2.09	0.30	527	10.36
55	-2.00	0.30	530	10.57
56	-1.91	0.31	533	10.78
57	-1.81	0.32	537	11.06
58	-1.71	0.32	540	11.34
59	-1.60	0.33	544	11.69
60	-1.49	0.35	548	12.11
61	-1.36	0.36	552	12.57
62	-1.23	0.38	557	13.13
63	-1.08	0.39	562	13.79
64	-0.91	0.42	568	14.60
65	-0.73	0.45	575	15.58
66	-0.51	0.48	582	16.77
67	-0.26	0.52	591	18.27
68	0.04	0.58	601	20.20
69	0.41	0.65	614	22.79
70	0.91	0.77	632	26.88
71	1.68	1.03	659	36.12
72	3.10	2.01	709	70.18

B.2: Primary

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-7.52	2.02	337	70.81
1	-6.06	1.05	388	36.68
2	-5.27	0.78	416	27.20
3	-4.75	0.67	434	23.28
4	-4.35	0.60	448	21.14
5	-4.01	0.57	460	19.81
6	-3.71	0.54	470	18.97
7	-3.42	0.53	480	18.41
8	-3.15	0.52	490	18.03
9	-2.89	0.51	499	17.82
10	-2.63	0.51	508	17.71
11	-2.37	0.51	517	17.75
12	-2.11	0.51	526	17.92
13	-1.85	0.52	535	18.24
14	-1.57	0.54	545	18.80
15	-1.27	0.56	556	19.64
16	-0.93	0.60	567	20.97
17	-0.54	0.66	581	23.14
18	-0.03	0.77	599	27.06
19	0.76	1.05	627	36.61
20	2.21	2.02	678	70.77

Strand 2: Writing Conventions				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.33	2.03	414	71.02
1	-3.85	1.06	465	36.93
2	-3.05	0.78	493	27.23
3	-2.54	0.66	511	23.07
4	-2.15	0.59	525	20.69
5	-1.83	0.55	536	19.15
6	-1.55	0.52	546	18.06
7	-1.29	0.50	555	17.33
8	-1.05	0.48	563	16.84
9	-0.83	0.47	571	16.56
10	-0.60	0.47	579	16.42
11	-0.38	0.47	587	16.45
12	-0.16	0.48	594	16.66
13	0.07	0.49	603	17.05
14	0.32	0.50	611	17.64
15	0.59	0.53	620	18.59
16	0.89	0.57	631	20.02
17	1.25	0.64	644	22.33
18	1.73	0.76	661	26.43
19	2.49	1.03	687	36.16
20	3.93	2.02	737	70.56

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.38	2.02	412	70.84
1	-3.91	1.05	463	36.72
2	-3.12	0.78	491	27.13
3	-2.61	0.66	509	23.10
4	-2.22	0.60	522	20.83
5	-1.89	0.55	534	19.36
6	-1.60	0.52	544	18.34
7	-1.34	0.50	553	17.64
8	-1.09	0.49	562	17.19
9	-0.86	0.48	570	16.91
10	-0.62	0.48	578	16.77
11	-0.39	0.48	586	16.77
12	-0.16	0.49	594	16.98
13	0.08	0.50	603	17.33
14	0.33	0.51	612	17.92
15	0.61	0.54	621	18.87
16	0.92	0.58	632	20.27
17	1.29	0.64	645	22.51
18	1.77	0.76	662	26.57
19	2.54	1.04	689	36.26
20	3.98	2.02	739	70.60

Strand 5: Speaking				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-6.28	2.02	380	70.67
1	-4.84	1.04	431	36.26
2	-4.07	0.75	458	26.39
3	-3.60	0.63	474	22.09
4	-3.25	0.56	486	19.50
5	-2.97	0.51	496	17.71
6	-2.73	0.47	504	16.45
7	-2.52	0.44	512	15.44
8	-2.34	0.42	518	14.67
9	-2.17	0.40	524	14.07
10	-2.01	0.39	530	13.62
11	-1.86	0.38	535	13.23
12	-1.72	0.37	540	12.95
13	-1.59	0.37	544	12.78
14	-1.46	0.36	549	12.64
15	-1.33	0.36	554	12.57
16	-1.20	0.36	558	12.53
17	-1.07	0.36	563	12.60
18	-0.94	0.36	567	12.71
19	-0.80	0.37	572	12.85
20	-0.67	0.37	577	13.09
21	-0.52	0.38	582	13.37
22	-0.37	0.39	587	13.76
23	-0.21	0.41	593	14.25
24	-0.04	0.42	599	14.84
25	0.15	0.45	605	15.58
26	0.36	0.47	613	16.52
27	0.60	0.51	621	17.78
28	0.88	0.56	631	19.46
29	1.23	0.63	643	21.98
30	1.70	0.75	659	26.25
31	2.45	1.03	686	36.09
32	3.89	2.02	736	70.56

Strand 10: Comprehension (Listening + Reading)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-7.63	2.02	333	70.67
1	-6.19	1.04	383	36.33
2	-5.42	0.76	410	26.64
3	-4.93	0.64	427	22.51
4	-4.56	0.58	440	20.16
5	-4.26	0.53	451	18.62
6	-3.99	0.50	460	17.50
7	-3.75	0.48	469	16.66
8	-3.53	0.46	476	16.03
9	-3.33	0.44	483	15.47
10	-3.14	0.43	490	15.05
11	-2.96	0.42	496	14.67
12	-2.79	0.41	502	14.35
13	-2.62	0.40	508	14.07
14	-2.46	0.40	514	13.83
15	-2.31	0.39	519	13.65
16	-2.16	0.39	524	13.48
17	-2.01	0.38	530	13.34
18	-1.87	0.38	535	13.23
19	-1.73	0.38	540	13.13
20	-1.59	0.37	544	13.09
21	-1.45	0.37	549	13.06
22	-1.31	0.37	554	13.06
23	-1.17	0.37	559	13.09
24	-1.03	0.38	564	13.13
25	-0.89	0.38	569	13.20
26	-0.74	0.38	574	13.34
27	-0.59	0.39	579	13.48
28	-0.44	0.39	584	13.69
29	-0.29	0.40	590	13.93
30	-0.12	0.41	596	14.28
31	0.05	0.42	602	14.67
32	0.23	0.43	608	15.19
33	0.43	0.45	615	15.82
34	0.64	0.48	622	16.66
35	0.88	0.51	631	17.82
36	1.16	0.55	641	19.39
37	1.51	0.62	653	21.84
38	1.97	0.75	669	26.08
39	2.72	1.03	695	35.91
40	4.15	2.01	745	70.46

Strand 11: Oral (Listening + Speaking)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-7.78	2.02	328	70.53
1	-6.35	1.03	378	36.05
2	-5.59	0.75	404	26.22
3	-5.13	0.63	421	21.98
4	-4.78	0.56	433	19.53
5	-4.49	0.51	443	17.89
6	-4.25	0.48	451	16.70
7	-4.03	0.45	459	15.79
8	-3.84	0.43	466	15.09
9	-3.66	0.41	472	14.46
10	-3.50	0.40	478	13.97
11	-3.34	0.39	483	13.51
12	-3.20	0.38	488	13.13
13	-3.06	0.37	493	12.81
14	-2.93	0.36	498	12.50
15	-2.80	0.35	502	12.22
16	-2.68	0.34	506	11.97
17	-2.57	0.34	510	11.73
18	-2.46	0.33	514	11.52
19	-2.35	0.32	518	11.34
20	-2.25	0.32	521	11.17
21	-2.15	0.32	525	11.03
22	-2.05	0.31	528	10.89
23	-1.95	0.31	532	10.78
24	-1.86	0.31	535	10.71
25	-1.76	0.30	538	10.64
26	-1.67	0.30	541	10.57
27	-1.58	0.30	545	10.54
28	-1.49	0.30	548	10.54
29	-1.40	0.30	551	10.54
30	-1.31	0.30	554	10.54
31	-1.22	0.30	557	10.61
32	-1.13	0.30	561	10.64
33	-1.03	0.31	564	10.75
34	-0.94	0.31	567	10.85
35	-0.84	0.31	571	10.99
36	-0.74	0.32	574	11.17
37	-0.63	0.32	578	11.34
38	-0.53	0.33	582	11.59
39	-0.42	0.34	585	11.87
40	-0.30	0.35	590	12.18
41	-0.17	0.36	594	12.57
42	-0.04	0.37	599	13.02

43	0.11	0.39	604	13.55
44	0.26	0.41	609	14.18
45	0.44	0.43	615	14.98
46	0.63	0.46	622	15.96
47	0.85	0.49	630	17.22
48	1.12	0.54	639	18.94
49	1.45	0.61	651	21.49
50	1.90	0.74	667	25.83
51	2.64	1.02	693	35.81
52	4.06	2.01	742	70.42

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-8.00	2.01	320	70.35
1	-6.59	1.02	369	35.70
2	-5.85	0.74	395	25.73
3	-5.41	0.61	411	21.39
4	-5.08	0.54	422	18.83
5	-4.82	0.49	431	17.12
6	-4.59	0.45	439	15.86
7	-4.40	0.43	446	14.88
8	-4.23	0.40	452	14.11
9	-4.07	0.38	457	13.44
10	-3.93	0.37	462	12.92
11	-3.80	0.36	467	12.46
12	-3.68	0.34	471	12.04
13	-3.56	0.33	475	11.69
14	-3.46	0.32	479	11.34
15	-3.35	0.32	483	11.06
16	-3.26	0.31	486	10.78
17	-3.16	0.30	489	10.54
18	-3.07	0.30	492	10.33
19	-2.99	0.29	495	10.12
20	-2.90	0.28	498	9.94
21	-2.83	0.28	501	9.77
22	-2.75	0.27	504	9.59
23	-2.67	0.27	506	9.45
24	-2.60	0.27	509	9.31
25	-2.53	0.26	511	9.17
26	-2.46	0.26	514	9.03
27	-2.40	0.26	516	8.93
28	-2.33	0.25	518	8.82
29	-2.27	0.25	521	8.72
30	-2.21	0.25	523	8.61
31	-2.15	0.24	525	8.54
32	-2.09	0.24	527	8.44
33	-2.03	0.24	529	8.37
34	-1.98	0.24	531	8.30
35	-1.92	0.24	533	8.23

36	-1.86	0.23	535	8.19
37	-1.81	0.23	537	8.12
38	-1.76	0.23	539	8.09
39	-1.70	0.23	540	8.02
40	-1.65	0.23	542	7.98
41	-1.60	0.23	544	7.95
42	-1.55	0.23	546	7.91
43	-1.50	0.23	548	7.88
44	-1.44	0.23	549	7.88
45	-1.39	0.22	551	7.84
46	-1.34	0.22	553	7.81
47	-1.29	0.22	555	7.81
48	-1.24	0.22	556	7.81
49	-1.19	0.22	558	7.81
50	-1.14	0.22	560	7.77
51	-1.09	0.22	562	7.77
52	-1.04	0.22	563	7.81
53	-1.00	0.22	565	7.81
54	-0.95	0.22	567	7.81
55	-0.90	0.22	569	7.81
56	-0.85	0.22	570	7.84
57	-0.79	0.22	572	7.84
58	-0.74	0.23	574	7.88
59	-0.69	0.23	576	7.91
60	-0.64	0.23	578	7.95
61	-0.59	0.23	579	7.98
62	-0.54	0.23	581	8.02
63	-0.48	0.23	583	8.05
64	-0.43	0.23	585	8.12
65	-0.38	0.23	587	8.16
66	-0.32	0.24	589	8.23
67	-0.27	0.24	591	8.26
68	-0.21	0.24	593	8.33
69	-0.15	0.24	595	8.40
70	-0.09	0.24	597	8.51
71	-0.03	0.25	599	8.58
72	0.03	0.25	601	8.68
73	0.09	0.25	603	8.79
74	0.15	0.25	605	8.89
75	0.22	0.26	608	9.00
76	0.29	0.26	610	9.10
77	0.36	0.26	612	9.24
78	0.43	0.27	615	9.38
79	0.50	0.27	617	9.56
80	0.58	0.28	620	9.73
81	0.65	0.28	623	9.91
82	0.74	0.29	626	10.12
83	0.82	0.30	629	10.36
84	0.91	0.30	632	10.61
85	1.01	0.31	635	10.89
86	1.11	0.32	639	11.20
87	1.21	0.33	642	11.59

88	1.33	0.34	646	11.97
89	1.45	0.36	651	12.46
90	1.58	0.37	655	12.99
91	1.73	0.39	660	13.62
92	1.89	0.41	666	14.39
93	2.07	0.44	672	15.30
94	2.27	0.47	680	16.45
95	2.51	0.51	688	17.89
96	2.80	0.57	698	19.85
97	3.17	0.65	711	22.65
98	3.67	0.78	728	27.16
99	4.47	1.06	757	37.07
100	5.96	2.03	809	71.19

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.89	2.03	394	71.12
1	-4.41	1.06	446	37.03
2	-3.60	0.78	474	27.34
3	-3.08	0.66	492	23.10
4	-2.70	0.59	506	20.62
5	-2.38	0.54	517	18.97
6	-2.10	0.51	526	17.78
7	-1.86	0.48	535	16.91
8	-1.63	0.46	543	16.24
9	-1.42	0.45	550	15.75
10	-1.23	0.44	557	15.40
11	-1.03	0.43	564	15.16
12	-0.85	0.43	570	14.98
13	-0.67	0.43	577	14.91
14	-0.48	0.43	583	14.95
15	-0.30	0.43	589	15.02
16	-0.11	0.44	596	15.23
17	0.08	0.44	603	15.51
18	0.28	0.45	610	15.86
19	0.49	0.47	617	16.38
20	0.72	0.49	625	17.05
21	0.97	0.51	634	17.92
22	1.25	0.55	644	19.08
23	1.57	0.59	655	20.62
24	1.95	0.65	668	22.68
25	2.42	0.73	685	25.59
26	3.05	0.86	707	30.07
27	4.00	1.13	740	39.41
28	5.59	2.07	796	72.38

B.3: Elementary

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.62	2.04	403	71.33
1	-4.12	1.08	456	37.77
2	-3.25	0.82	486	28.84
3	-2.66	0.73	507	25.38
4	-2.17	0.67	524	23.49
5	-1.75	0.63	539	22.19
6	-1.36	0.60	552	21.14
7	-1.01	0.58	565	20.27
8	-0.69	0.56	576	19.53
9	-0.39	0.54	586	18.97
10	-0.10	0.53	597	18.62
11	0.18	0.53	606	18.48
12	0.46	0.53	616	18.52
13	0.74	0.54	626	18.80
14	1.04	0.55	636	19.36
15	1.36	0.58	648	20.30
16	1.72	0.62	660	21.74
17	2.15	0.69	675	24.01
18	2.69	0.80	694	28.00
19	3.53	1.07	723	37.42
20	5.02	2.04	776	71.26

Strand 2: Writing Conventions				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.14	2.04	420	71.40
1	-3.63	1.08	473	37.66
2	-2.78	0.81	503	28.28
3	-2.22	0.69	522	24.29
4	-1.79	0.63	537	22.02
5	-1.42	0.59	550	20.51
6	-1.10	0.56	562	19.46
7	-0.80	0.54	572	18.73
8	-0.52	0.52	582	18.20
9	-0.26	0.51	591	17.82
10	0.00	0.50	600	17.61
11	0.25	0.50	609	17.54
12	0.50	0.50	618	17.61
13	0.76	0.51	627	17.89
14	1.03	0.53	636	18.38
15	1.32	0.55	646	19.22
16	1.64	0.59	657	20.55
17	2.02	0.65	671	22.75
18	2.51	0.76	688	26.74
19	3.29	1.04	715	36.33
20	4.73	2.02	766	70.63

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.67	2.06	437	72.14
1	-3.09	1.13	492	39.52
2	-2.11	0.89	526	31.15
3	-1.41	0.78	551	27.34
4	-0.87	0.70	570	24.47
5	-0.42	0.63	585	22.16
6	-0.06	0.58	598	20.41
7	0.26	0.55	609	19.15
8	0.55	0.52	619	18.27
9	0.81	0.51	628	17.68
10	1.06	0.50	637	17.33
11	1.30	0.49	646	17.19
12	1.55	0.49	654	17.26
13	1.79	0.50	663	17.57
14	2.05	0.52	672	18.10
15	2.33	0.54	682	19.01
16	2.65	0.58	693	20.37
17	3.02	0.65	706	22.65
18	3.51	0.76	723	26.71
19	4.29	1.04	750	36.37
20	5.73	2.02	801	70.67

Strand 5: Speaking				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-6.04	2.02	389	70.74
1	-4.59	1.04	439	36.37
2	-3.82	0.76	466	26.57
3	-3.34	0.64	483	22.33
4	-2.98	0.57	496	19.81
5	-2.68	0.52	506	18.13
6	-2.43	0.48	515	16.94
7	-2.21	0.46	523	16.03
8	-2.01	0.44	530	15.30
9	-1.83	0.42	536	14.70
10	-1.66	0.41	542	14.21
11	-1.50	0.39	548	13.79
12	-1.35	0.39	553	13.48
13	-1.20	0.38	558	13.20
14	-1.06	0.37	563	12.99
15	-0.92	0.37	568	12.81
16	-0.79	0.36	572	12.74
17	-0.66	0.36	577	12.71
18	-0.53	0.36	582	12.74
19	-0.39	0.37	586	12.85
20	-0.26	0.37	591	13.02
21	-0.12	0.38	596	13.27
22	0.03	0.39	601	13.62
23	0.19	0.40	607	14.04
24	0.36	0.42	612	14.60
25	0.54	0.44	619	15.33
26	0.74	0.46	626	16.24
27	0.97	0.50	634	17.47
28	1.25	0.55	644	19.15
29	1.58	0.62	655	21.67
30	2.04	0.74	671	25.97
31	2.79	1.03	697	35.88
32	4.21	2.01	747	70.42

Strand 10: Comprehension (Listening + Reading)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.97	2.02	391	70.84
1	-4.51	1.05	442	36.72
2	-3.71	0.78	470	27.27
3	-3.19	0.67	488	23.35
4	-2.79	0.61	502	21.18
5	-2.45	0.57	514	19.78
6	-2.15	0.54	525	18.76
7	-1.87	0.51	534	17.99
8	-1.62	0.50	543	17.33
9	-1.38	0.48	552	16.70
10	-1.16	0.46	559	16.17
11	-0.95	0.45	567	15.65
12	-0.76	0.43	573	15.19
13	-0.58	0.42	580	14.77
14	-0.40	0.41	586	14.39
15	-0.24	0.40	592	14.07
16	-0.08	0.39	597	13.79
17	0.07	0.39	603	13.55
18	0.22	0.38	608	13.34
19	0.36	0.38	613	13.16
20	0.50	0.37	618	13.06
21	0.64	0.37	623	12.95
22	0.78	0.37	627	12.92
23	0.92	0.37	632	12.88
24	1.05	0.37	637	12.92
25	1.19	0.37	642	12.99
26	1.33	0.37	646	13.09
27	1.47	0.38	651	13.23
28	1.61	0.38	657	13.41
29	1.77	0.39	662	13.69
30	1.92	0.40	667	14.00
31	2.09	0.41	673	14.42
32	2.26	0.43	679	14.95
33	2.45	0.45	686	15.61
34	2.66	0.47	693	16.49
35	2.90	0.50	701	17.64
36	3.18	0.55	711	19.25
37	3.52	0.62	723	21.74
38	3.98	0.74	739	26.01
39	4.72	1.03	765	35.88
40	6.14	2.01	815	70.42

Strand 11: Oral (Listening + Speaking)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-6.56	2.01	370	70.46
1	-5.13	1.03	420	35.95
2	-4.38	0.74	447	26.04
3	-3.92	0.62	463	21.77
4	-3.58	0.55	475	19.29
5	-3.31	0.50	484	17.61
6	-3.07	0.47	493	16.42
7	-2.86	0.44	500	15.47
8	-2.68	0.42	506	14.74
9	-2.50	0.40	512	14.14
10	-2.35	0.39	518	13.65
11	-2.20	0.38	523	13.23
12	-2.06	0.37	528	12.85
13	-1.93	0.36	532	12.53
14	-1.80	0.35	537	12.25
15	-1.68	0.34	541	11.97
16	-1.57	0.34	545	11.76
17	-1.46	0.33	549	11.55
18	-1.35	0.32	553	11.34
19	-1.25	0.32	556	11.20
20	-1.15	0.32	560	11.06
21	-1.05	0.31	563	10.92
22	-0.95	0.31	567	10.82
23	-0.85	0.31	570	10.75
24	-0.76	0.31	573	10.68
25	-0.67	0.30	577	10.64
26	-0.58	0.30	580	10.61
27	-0.48	0.30	583	10.61
28	-0.39	0.30	586	10.64
29	-0.30	0.31	590	10.68
30	-0.20	0.31	593	10.71
31	-0.11	0.31	596	10.82
32	-0.01	0.31	600	10.92
33	0.09	0.32	603	11.03
34	0.19	0.32	607	11.17
35	0.29	0.32	610	11.34
36	0.40	0.33	614	11.52
37	0.51	0.34	618	11.76
38	0.62	0.34	622	12.01
39	0.74	0.35	626	12.29
40	0.87	0.36	630	12.64
41	1.00	0.37	635	13.02
42	1.15	0.39	640	13.48
43	1.30	0.40	646	14.00

44	1.47	0.42	651	14.63
45	1.65	0.44	658	15.44
46	1.86	0.47	665	16.42
47	2.10	0.51	673	17.71
48	2.38	0.56	683	19.43
49	2.73	0.63	695	22.02
50	3.20	0.75	712	26.32
51	3.96	1.03	739	36.19
52	5.40	2.02	789	70.63

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-6.93	2.01	357	70.28
1	-5.52	1.02	407	35.56
2	-4.80	0.73	432	25.55
3	-4.36	0.61	447	21.18
4	-4.04	0.53	459	18.59
5	-3.78	0.48	468	16.84
6	-3.57	0.45	475	15.58
7	-3.38	0.42	482	14.60
8	-3.22	0.40	487	13.83
9	-3.07	0.38	493	13.20
10	-2.93	0.36	497	12.64
11	-2.81	0.35	502	12.18
12	-2.69	0.34	506	11.80
13	-2.58	0.33	510	11.45
14	-2.47	0.32	513	11.13
15	-2.37	0.31	517	10.85
16	-2.28	0.30	520	10.61
17	-2.19	0.30	523	10.40
18	-2.10	0.29	526	10.19
19	-2.02	0.29	529	10.01
20	-1.94	0.28	532	9.84
21	-1.86	0.28	535	9.70
22	-1.79	0.27	537	9.52
23	-1.71	0.27	540	9.42
24	-1.64	0.27	543	9.28
25	-1.57	0.26	545	9.17
26	-1.50	0.26	547	9.07
27	-1.44	0.26	550	8.96
28	-1.37	0.25	552	8.89
29	-1.31	0.25	554	8.79
30	-1.25	0.25	556	8.72
31	-1.18	0.25	559	8.65
32	-1.12	0.25	561	8.58
33	-1.06	0.24	563	8.51
34	-1.00	0.24	565	8.47
35	-0.95	0.24	567	8.40
36	-0.89	0.24	569	8.37

37	-0.83	0.24	571	8.33
38	-0.77	0.24	573	8.30
39	-0.72	0.24	575	8.26
40	-0.66	0.24	577	8.23
41	-0.61	0.23	579	8.19
42	-0.55	0.23	581	8.19
43	-0.50	0.23	583	8.16
44	-0.44	0.23	584	8.16
45	-0.39	0.23	586	8.12
46	-0.33	0.23	588	8.12
47	-0.28	0.23	590	8.12
48	-0.23	0.23	592	8.12
49	-0.17	0.23	594	8.12
50	-0.12	0.23	596	8.12
51	-0.06	0.23	598	8.16
52	-0.01	0.23	600	8.16
53	0.05	0.23	602	8.16
54	0.10	0.23	604	8.19
55	0.16	0.23	605	8.19
56	0.21	0.24	607	8.23
57	0.27	0.24	609	8.26
58	0.32	0.24	611	8.26
59	0.38	0.24	613	8.30
60	0.43	0.24	615	8.33
61	0.49	0.24	617	8.37
62	0.55	0.24	619	8.40
63	0.61	0.24	621	8.44
64	0.67	0.24	623	8.51
65	0.73	0.24	625	8.54
66	0.79	0.25	628	8.58
67	0.85	0.25	630	8.65
68	0.91	0.25	632	8.72
69	0.97	0.25	634	8.75
70	1.03	0.25	636	8.82
71	1.10	0.26	638	8.93
72	1.16	0.26	641	9.00
73	1.23	0.26	643	9.07
74	1.30	0.26	645	9.17
75	1.37	0.27	648	9.28
76	1.44	0.27	650	9.38
77	1.51	0.27	653	9.49
78	1.59	0.28	656	9.63
79	1.67	0.28	658	9.77
80	1.74	0.28	661	9.94
81	1.83	0.29	664	10.12
82	1.91	0.29	667	10.29
83	2.00	0.30	670	10.50
84	2.09	0.31	673	10.75
85	2.19	0.31	677	10.99
86	2.29	0.32	680	11.31
87	2.40	0.33	684	11.66
88	2.51	0.34	688	12.01

89	2.64	0.36	692	12.46
90	2.77	0.37	697	12.99
91	2.91	0.39	702	13.58
92	3.07	0.41	708	14.32
93	3.25	0.44	714	15.23
94	3.45	0.47	721	16.38
95	3.69	0.51	729	17.92
96	3.99	0.57	740	19.99
97	4.36	0.66	753	23.10
98	4.89	0.81	771	28.25
99	5.78	1.11	802	38.89
100	7.37	2.07	858	72.59

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.30	2.03	414	70.91
1	-3.84	1.05	466	36.65
2	-3.06	0.77	493	26.78
3	-2.57	0.64	510	22.47
4	-2.20	0.57	523	20.02
5	-1.90	0.53	533	18.45
6	-1.64	0.50	543	17.43
7	-1.40	0.48	551	16.73
8	-1.18	0.46	559	16.24
9	-0.97	0.46	566	15.93
10	-0.76	0.45	573	15.72
11	-0.56	0.45	580	15.58
12	-0.37	0.44	587	15.51
13	-0.17	0.44	594	15.47
14	0.03	0.44	601	15.51
15	0.22	0.45	608	15.58
16	0.42	0.45	615	15.72
17	0.63	0.45	622	15.89
18	0.84	0.46	629	16.21
19	1.06	0.48	637	16.63
20	1.29	0.49	645	17.19
21	1.54	0.51	654	17.96
22	1.82	0.54	664	19.01
23	2.14	0.59	675	20.51
24	2.52	0.65	688	22.75
25	3.00	0.75	705	26.32
26	3.71	0.94	730	32.87
27	4.94	1.31	773	45.75
28	6.89	2.18	841	76.13

B.4: Middle Grades

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.51	2.05	442	71.75
1	-2.97	1.09	496	38.26
2	-2.09	0.82	527	28.84
3	-1.51	0.71	547	24.68
4	-1.07	0.64	563	22.23
5	-0.69	0.59	576	20.55
6	-0.37	0.56	587	19.43
7	-0.07	0.53	597	18.62
8	0.20	0.52	607	18.06
9	0.46	0.51	616	17.71
10	0.72	0.50	625	17.57
11	0.97	0.50	634	17.57
12	1.22	0.51	643	17.71
13	1.48	0.52	652	18.06
14	1.76	0.53	662	18.66
15	2.05	0.56	672	19.53
16	2.39	0.60	684	20.90
17	2.78	0.66	697	23.10
18	3.29	0.77	715	27.06
19	4.08	1.05	743	36.61
20	5.53	2.02	794	70.77

Strand 2: Writing Conventions				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.38	2.02	447	70.77
1	-2.93	1.04	498	36.51
2	-2.14	0.77	525	26.81
3	-1.65	0.65	542	22.72
4	-1.27	0.58	555	20.37
5	-0.96	0.54	566	18.83
6	-0.69	0.51	576	17.78
7	-0.44	0.49	585	17.01
8	-0.21	0.47	593	16.49
9	0.00	0.46	600	16.10
10	0.21	0.45	607	15.82
11	0.41	0.45	615	15.68
12	0.61	0.45	622	15.61
13	0.81	0.45	629	15.65
14	1.02	0.45	636	15.79
15	1.22	0.46	643	16.00
16	1.44	0.47	650	16.38
17	1.66	0.48	658	16.87
18	1.90	0.50	667	17.61
19	2.17	0.53	676	18.62
20	2.48	0.57	687	20.09
21	2.84	0.64	700	22.44
22	3.33	0.76	716	26.53
23	4.09	1.04	743	36.26
24	5.53	2.02	794	70.60

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-3.79	2.02	467	70.74
1	-2.34	1.04	518	36.44
2	-1.56	0.76	545	26.71
3	-1.07	0.65	562	22.58
4	-0.70	0.58	575	20.20
5	-0.40	0.53	586	18.66
6	-0.13	0.50	596	17.61
7	0.11	0.48	604	16.80
8	0.34	0.46	612	16.24
9	0.55	0.45	619	15.82
10	0.75	0.44	626	15.54
11	0.94	0.44	633	15.40
12	1.14	0.44	640	15.33
13	1.33	0.44	646	15.33
14	1.52	0.44	653	15.47
15	1.72	0.45	660	15.72
16	1.93	0.46	667	16.10
17	2.15	0.48	675	16.63
18	2.38	0.50	683	17.40
19	2.64	0.53	693	18.45
20	2.94	0.57	703	19.99
21	3.31	0.64	716	22.37
22	3.79	0.76	733	26.57
23	4.56	1.04	760	36.30
24	6.00	2.02	810	70.67

Strand 5: Speaking				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.34	2.01	448	70.49
1	-2.91	1.03	498	35.95
2	-2.16	0.74	524	25.97
3	-1.70	0.62	540	21.60
4	-1.37	0.54	552	19.01
5	-1.10	0.49	561	17.26
6	-0.88	0.46	569	16.00
7	-0.68	0.43	576	15.05
8	-0.50	0.41	582	14.35
9	-0.34	0.39	588	13.76
10	-0.19	0.38	593	13.34
11	-0.05	0.37	598	13.02
12	0.09	0.37	603	12.78
13	0.22	0.36	608	12.60
14	0.35	0.36	612	12.46
15	0.47	0.36	617	12.43
16	0.60	0.36	621	12.43
17	0.73	0.36	625	12.50
18	0.85	0.36	630	12.64
19	0.99	0.37	635	12.81
20	1.12	0.37	639	13.02
21	1.27	0.38	644	13.34
22	1.41	0.39	650	13.69
23	1.57	0.40	655	14.14
24	1.74	0.42	661	14.70
25	1.93	0.44	667	15.40
26	2.13	0.47	675	16.28
27	2.37	0.50	683	17.47
28	2.64	0.55	692	19.11
29	2.97	0.62	704	21.60
30	3.43	0.74	720	25.87
31	4.17	1.02	746	35.77
32	5.59	2.01	795	70.39

Strand 10: Comprehension (Listening + Reading)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.92	2.03	428	70.88
1	-3.46	1.05	479	36.58
2	-2.68	0.76	506	26.74
3	-2.19	0.64	523	22.44
4	-1.83	0.57	536	19.85
5	-1.54	0.52	546	18.10
6	-1.29	0.48	555	16.80
7	-1.07	0.45	563	15.82
8	-0.88	0.43	569	15.05
9	-0.70	0.41	576	14.42
10	-0.54	0.40	581	13.90
11	-0.38	0.38	587	13.44
12	-0.24	0.37	592	13.09
13	-0.10	0.37	596	12.78
14	0.03	0.36	601	12.50
15	0.15	0.35	605	12.29
16	0.28	0.35	610	12.11
17	0.39	0.34	614	11.94
18	0.51	0.34	618	11.83
19	0.62	0.34	622	11.73
20	0.73	0.33	626	11.66
21	0.84	0.33	630	11.59
22	0.95	0.33	633	11.55
23	1.06	0.33	637	11.55
24	1.17	0.33	641	11.59
25	1.28	0.33	645	11.62
26	1.39	0.33	649	11.66
27	1.51	0.34	653	11.76
28	1.62	0.34	657	11.87
29	1.74	0.34	661	12.01
30	1.86	0.35	665	12.18
31	1.98	0.35	669	12.39
32	2.11	0.36	674	12.67
33	2.24	0.37	678	12.99
34	2.38	0.38	683	13.37
35	2.54	0.40	689	13.83
36	2.70	0.41	694	14.39
37	2.88	0.43	701	15.12
38	3.07	0.46	708	16.03
39	3.30	0.49	715	17.26
40	3.57	0.54	725	18.94
41	3.90	0.61	736	21.46
42	4.34	0.74	752	25.76
43	5.08	1.02	778	35.74
44	6.50	2.01	827	70.35

Strand 11: Oral (Listening + Speaking)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.13	2.02	420	70.63
1	-3.70	1.03	471	36.16
2	-2.94	0.75	497	26.22
3	-2.47	0.63	513	21.88
4	-2.13	0.55	525	19.25
5	-1.85	0.50	535	17.47
6	-1.62	0.46	543	16.14
7	-1.42	0.43	550	15.12
8	-1.25	0.41	556	14.28
9	-1.09	0.39	562	13.62
10	-0.94	0.37	567	13.06
11	-0.81	0.36	572	12.57
12	-0.68	0.35	576	12.15
13	-0.57	0.34	580	11.80
14	-0.46	0.33	584	11.48
15	-0.35	0.32	588	11.24
16	-0.25	0.32	591	11.03
17	-0.15	0.31	595	10.82
18	-0.06	0.30	598	10.64
19	0.03	0.30	601	10.54
20	0.12	0.30	604	10.40
21	0.21	0.30	607	10.33
22	0.30	0.29	610	10.26
23	0.38	0.29	613	10.22
24	0.47	0.29	616	10.19
25	0.55	0.29	619	10.15
26	0.64	0.29	622	10.15
27	0.72	0.29	625	10.19
28	0.81	0.29	628	10.22
29	0.89	0.29	631	10.26
30	0.98	0.30	634	10.33
31	1.07	0.30	637	10.43
32	1.16	0.30	641	10.50
33	1.25	0.30	644	10.64
34	1.34	0.31	647	10.78
35	1.44	0.31	650	10.92
36	1.54	0.32	654	11.10
37	1.64	0.32	657	11.31
38	1.75	0.33	661	11.55
39	1.86	0.34	665	11.83
40	1.98	0.35	669	12.11
41	2.10	0.36	674	12.50
42	2.23	0.37	678	12.92
43	2.37	0.38	683	13.41
44	2.53	0.40	688	14.00

45	2.70	0.42	694	14.77
46	2.89	0.45	701	15.72
47	3.10	0.48	709	16.94
48	3.36	0.53	718	18.66
49	3.68	0.61	729	21.21
50	4.12	0.73	744	25.55
51	4.85	1.02	770	35.56
52	6.26	2.01	819	70.28

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.71	2.01	400	70.25
1	-4.30	1.01	449	35.49
2	-3.58	0.73	475	25.38
3	-3.15	0.60	490	20.93
4	-2.84	0.52	501	18.31
5	-2.59	0.47	509	16.49
6	-2.39	0.43	516	15.16
7	-2.21	0.40	523	14.14
8	-2.06	0.38	528	13.30
9	-1.92	0.36	533	12.64
10	-1.80	0.34	537	12.04
11	-1.68	0.33	541	11.55
12	-1.58	0.32	545	11.13
13	-1.48	0.31	548	10.75
14	-1.39	0.30	551	10.40
15	-1.30	0.29	554	10.12
16	-1.22	0.28	557	9.84
17	-1.14	0.28	560	9.63
18	-1.07	0.27	563	9.38
19	-1.00	0.26	565	9.21
20	-0.93	0.26	567	9.03
21	-0.87	0.25	570	8.86
22	-0.80	0.25	572	8.72
23	-0.74	0.25	574	8.58
24	-0.68	0.24	576	8.47
25	-0.62	0.24	578	8.37
26	-0.57	0.24	580	8.26
27	-0.51	0.23	582	8.16
28	-0.46	0.23	584	8.09
29	-0.40	0.23	586	8.02
30	-0.35	0.23	588	7.95
31	-0.30	0.23	589	7.88
32	-0.25	0.22	591	7.81
33	-0.20	0.22	593	7.74
34	-0.15	0.22	595	7.70
35	-0.10	0.22	596	7.67
36	-0.06	0.22	598	7.60

37	-0.01	0.22	600	7.56
38	0.04	0.22	601	7.53
39	0.08	0.21	603	7.49
40	0.13	0.21	605	7.46
41	0.18	0.21	606	7.46
42	0.22	0.21	608	7.42
43	0.27	0.21	609	7.39
44	0.31	0.21	611	7.39
45	0.35	0.21	612	7.35
46	0.40	0.21	614	7.35
47	0.44	0.21	616	7.35
48	0.49	0.21	617	7.35
49	0.53	0.21	619	7.32
50	0.58	0.21	620	7.32
51	0.62	0.21	622	7.32
52	0.66	0.21	623	7.32
53	0.71	0.21	625	7.32
54	0.75	0.21	626	7.32
55	0.80	0.21	628	7.35
56	0.84	0.21	629	7.35
57	0.88	0.21	631	7.35
58	0.93	0.21	633	7.39
59	0.97	0.21	634	7.39
60	1.02	0.21	636	7.42
61	1.06	0.21	637	7.42
62	1.11	0.21	639	7.46
63	1.15	0.21	640	7.49
64	1.20	0.22	642	7.53
65	1.25	0.22	644	7.56
66	1.29	0.22	645	7.60
67	1.34	0.22	647	7.63
68	1.39	0.22	649	7.67
69	1.44	0.22	650	7.70
70	1.49	0.22	652	7.74
71	1.54	0.22	654	7.81
72	1.59	0.22	656	7.84
73	1.64	0.23	657	7.91
74	1.69	0.23	659	7.98
75	1.74	0.23	661	8.02
76	1.79	0.23	663	8.09
77	1.85	0.23	665	8.19
78	1.90	0.24	667	8.26
79	1.96	0.24	669	8.33
80	2.02	0.24	671	8.44
81	2.08	0.24	673	8.54
82	2.14	0.25	675	8.65
83	2.20	0.25	677	8.75
84	2.26	0.25	679	8.89
85	2.33	0.26	681	9.03
86	2.40	0.26	684	9.17
87	2.47	0.27	686	9.35
88	2.54	0.27	689	9.52

89	2.61	0.28	691	9.70
90	2.69	0.28	694	9.91
91	2.77	0.29	697	10.15
92	2.86	0.30	700	10.40
93	2.95	0.31	703	10.71
94	3.05	0.32	707	11.03
95	3.15	0.33	710	11.41
96	3.26	0.34	714	11.83
97	3.38	0.35	718	12.32
98	3.51	0.37	723	12.92
99	3.66	0.39	728	13.58
100	3.81	0.41	734	14.39
101	4.00	0.44	740	15.40
102	4.21	0.48	747	16.63
103	4.45	0.52	756	18.24
104	4.76	0.58	767	20.41
105	5.15	0.67	780	23.42
106	5.68	0.80	799	28.11
107	6.54	1.08	829	37.91
108	8.06	2.05	882	71.61

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.44	2.01	444	70.42
1	-3.02	1.02	494	35.77
2	-2.29	0.73	520	25.66
3	-1.84	0.61	535	21.25
4	-1.52	0.53	547	18.69
5	-1.26	0.49	556	17.01
6	-1.04	0.45	564	15.89
7	-0.85	0.43	570	15.16
8	-0.66	0.42	577	14.63
9	-0.49	0.41	583	14.28
10	-0.33	0.40	588	14.07
11	-0.17	0.40	594	14.00
12	-0.01	0.40	600	13.97
13	0.15	0.40	605	14.00
14	0.31	0.40	611	14.11
15	0.48	0.41	617	14.21
16	0.64	0.41	622	14.35
17	0.81	0.41	628	14.49
18	0.99	0.42	635	14.67
19	1.17	0.43	641	14.91
20	1.35	0.43	647	15.19
21	1.54	0.44	654	15.54
22	1.75	0.46	661	15.96
23	1.96	0.47	669	16.52
24	2.19	0.49	677	17.22
25	2.45	0.52	686	18.20
26	2.74	0.56	696	19.50
27	3.08	0.61	708	21.32
28	3.49	0.69	722	24.05
29	4.05	0.80	742	28.14
30	4.83	0.97	769	33.78
31	5.97	1.20	809	42.14
32	7.68	2.10	869	73.33

B.5: High School

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-3.77	2.03	468	70.88
1	-2.30	1.05	519	36.75
2	-1.51	0.78	547	27.20
3	-1.00	0.66	565	23.17
4	-0.60	0.60	579	20.93
5	-0.27	0.56	591	19.46
6	0.02	0.53	601	18.52
7	0.29	0.51	610	17.85
8	0.55	0.50	619	17.47
9	0.79	0.49	628	17.22
10	1.03	0.49	636	17.15
11	1.28	0.49	645	17.22
12	1.52	0.50	653	17.47
13	1.78	0.51	662	17.89
14	2.05	0.53	672	18.52
15	2.34	0.56	682	19.43
16	2.67	0.60	693	20.83
17	3.06	0.66	707	23.07
18	3.57	0.77	725	27.06
19	4.36	1.05	752	36.61
20	5.81	2.02	803	70.81

Strand 2: Writing Conventions				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.08	2.03	457	70.98
1	-2.61	1.06	509	36.96
2	-1.80	0.79	537	27.51
3	-1.27	0.67	555	23.56
4	-0.87	0.61	570	21.32
5	-0.52	0.57	582	19.88
6	-0.21	0.54	593	18.87
7	0.07	0.52	602	18.13
8	0.33	0.50	611	17.54
9	0.57	0.49	620	17.08
10	0.80	0.48	628	16.77
11	1.03	0.47	636	16.49
12	1.25	0.47	644	16.35
13	1.47	0.47	651	16.28
14	1.68	0.47	659	16.31
15	1.90	0.47	667	16.42
16	2.13	0.48	674	16.70
17	2.36	0.49	683	17.12
18	2.61	0.51	691	17.78
19	2.88	0.54	701	18.73
20	3.19	0.58	712	20.16
21	3.56	0.64	725	22.47
22	4.04	0.76	741	26.53
23	4.81	1.04	768	36.23
24	6.25	2.02	819	70.60

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.14	2.10	455	73.47
1	-2.47	1.16	514	40.50
2	-1.49	0.85	548	29.86
3	-0.89	0.71	569	24.68
4	-0.46	0.62	584	21.67
5	-0.11	0.56	596	19.71
6	0.18	0.53	606	18.38
7	0.44	0.50	616	17.40
8	0.68	0.48	624	16.73
9	0.90	0.46	632	16.24
10	1.11	0.45	639	15.89
11	1.32	0.45	646	15.68
12	1.52	0.45	653	15.58
13	1.72	0.45	660	15.58
14	1.91	0.45	667	15.68
15	2.12	0.46	674	15.93
16	2.33	0.47	682	16.28
17	2.55	0.48	689	16.77
18	2.79	0.50	698	17.50
19	3.06	0.53	707	18.52
20	3.36	0.57	718	20.02
21	3.72	0.64	730	22.37
22	4.20	0.76	747	26.50
23	4.97	1.03	774	36.19
24	6.41	2.02	824	70.60

Strand 5: Speaking				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-3.55	2.01	476	70.35
1	-2.13	1.02	526	35.70
2	-1.39	0.73	551	25.69
3	-0.95	0.61	567	21.32
4	-0.62	0.54	578	18.73
5	-0.36	0.49	587	17.01
6	-0.15	0.45	595	15.75
7	0.05	0.42	602	14.81
8	0.22	0.40	608	14.07
9	0.37	0.39	613	13.51
10	0.51	0.37	618	13.02
11	0.65	0.36	623	12.67
12	0.78	0.35	627	12.39
13	0.90	0.35	632	12.18
14	1.02	0.34	636	12.04
15	1.14	0.34	640	11.94
16	1.25	0.34	644	11.90
17	1.37	0.34	648	11.94
18	1.49	0.34	652	12.01
19	1.61	0.35	656	12.15
20	1.73	0.35	661	12.36
21	1.86	0.36	665	12.64
22	1.99	0.37	670	13.02
23	2.14	0.39	675	13.48
24	2.29	0.40	680	14.04
25	2.46	0.42	686	14.81
26	2.65	0.45	693	15.75
27	2.87	0.49	700	17.01
28	3.13	0.54	709	18.73
29	3.45	0.61	721	21.32
30	3.90	0.73	736	25.69
31	4.63	1.02	762	35.67
32	6.05	2.01	812	70.35

Strand 10: Comprehension (Listening + Reading)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.69	2.04	436	71.37
1	-3.20	1.07	488	37.31
2	-2.38	0.78	517	27.34
3	-1.87	0.65	534	22.82
4	-1.50	0.57	547	20.09
5	-1.20	0.52	558	18.24
6	-0.95	0.48	567	16.91
7	-0.73	0.45	574	15.89
8	-0.53	0.43	581	15.09
9	-0.36	0.41	588	14.42
10	-0.19	0.40	593	13.90
11	-0.04	0.38	599	13.44
12	0.10	0.37	604	13.09
13	0.24	0.37	608	12.78
14	0.37	0.36	613	12.50
15	0.50	0.35	617	12.29
16	0.62	0.35	622	12.11
17	0.74	0.34	626	11.97
18	0.85	0.34	630	11.83
19	0.97	0.34	634	11.73
20	1.08	0.33	638	11.66
21	1.19	0.33	642	11.62
22	1.30	0.33	645	11.59
23	1.41	0.33	649	11.59
24	1.52	0.33	653	11.62
25	1.63	0.33	657	11.66
26	1.74	0.34	661	11.73
27	1.85	0.34	665	11.80
28	1.97	0.34	669	11.94
29	2.09	0.35	673	12.08
30	2.21	0.35	677	12.25
31	2.33	0.36	682	12.46
32	2.46	0.36	686	12.74
33	2.60	0.37	691	13.02
34	2.74	0.38	696	13.41
35	2.89	0.40	701	13.86
36	3.06	0.41	707	14.46
37	3.24	0.43	713	15.16
38	3.44	0.46	720	16.07
39	3.66	0.49	728	17.26
40	3.93	0.54	737	18.94
41	4.26	0.61	749	21.42
42	4.70	0.74	765	25.73
43	5.44	1.02	790	35.70
44	6.86	2.01	840	70.35

Strand 11: Oral (Listening + Speaking)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.37	2.01	447	70.35
1	-2.95	1.02	497	35.67
2	-2.22	0.73	522	25.69
3	-1.77	0.61	538	21.32
4	-1.45	0.54	549	18.73
5	-1.19	0.49	558	16.98
6	-0.97	0.45	566	15.72
7	-0.78	0.42	573	14.74
8	-0.61	0.40	578	13.93
9	-0.46	0.38	584	13.30
10	-0.32	0.36	589	12.74
11	-0.20	0.35	593	12.29
12	-0.08	0.34	597	11.90
13	0.04	0.33	601	11.59
14	0.14	0.32	605	11.27
15	0.25	0.32	609	11.03
16	0.34	0.31	612	10.78
17	0.44	0.30	615	10.61
18	0.53	0.30	618	10.43
19	0.61	0.29	622	10.29
20	0.70	0.29	624	10.15
21	0.78	0.29	627	10.05
22	0.87	0.29	630	9.98
23	0.95	0.28	633	9.91
24	1.03	0.28	636	9.84
25	1.10	0.28	639	9.80
26	1.18	0.28	641	9.80
27	1.26	0.28	644	9.80
28	1.34	0.28	647	9.80
29	1.42	0.28	650	9.84
30	1.50	0.28	652	9.91
31	1.58	0.29	655	9.98
32	1.66	0.29	658	10.05
33	1.74	0.29	661	10.15
34	1.83	0.29	664	10.29
35	1.92	0.30	667	10.47
36	2.01	0.30	670	10.64
37	2.10	0.31	674	10.85
38	2.20	0.32	677	11.10
39	2.31	0.33	681	11.41
40	2.42	0.34	685	11.73
41	2.53	0.35	689	12.11
42	2.66	0.36	693	12.57
43	2.79	0.37	698	13.09
44	2.94	0.39	703	13.72

45	3.10	0.42	709	14.53
46	3.29	0.44	715	15.51
47	3.50	0.48	722	16.77
48	3.75	0.53	731	18.52
49	4.07	0.60	742	21.11
50	4.50	0.73	758	25.48
51	5.23	1.02	783	35.53
52	6.64	2.01	832	70.25

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-5.36	2.01	412	70.46
1	-3.94	1.02	462	35.84
2	-3.20	0.74	488	25.80
3	-2.75	0.61	504	21.35
4	-2.43	0.53	515	18.69
5	-2.17	0.48	524	16.87
6	-1.95	0.44	532	15.54
7	-1.77	0.41	538	14.49
8	-1.61	0.39	544	13.65
9	-1.46	0.37	549	12.99
10	-1.33	0.35	553	12.39
11	-1.21	0.34	558	11.90
12	-1.10	0.33	562	11.45
13	-1.00	0.32	565	11.06
14	-0.90	0.31	569	10.71
15	-0.81	0.30	572	10.43
16	-0.72	0.29	575	10.15
17	-0.64	0.28	578	9.91
18	-0.56	0.28	580	9.66
19	-0.49	0.27	583	9.45
20	-0.41	0.27	586	9.28
21	-0.35	0.26	588	9.10
22	-0.28	0.26	590	8.93
23	-0.21	0.25	592	8.79
24	-0.15	0.25	595	8.65
25	-0.09	0.24	597	8.51
26	-0.03	0.24	599	8.40
27	0.02	0.24	601	8.30
28	0.08	0.23	603	8.19
29	0.13	0.23	605	8.09
30	0.19	0.23	607	7.98
31	0.24	0.23	608	7.91
32	0.29	0.22	610	7.84
33	0.34	0.22	612	7.77
34	0.39	0.22	614	7.70
35	0.44	0.22	615	7.63
36	0.48	0.22	617	7.60

37	0.53	0.22	619	7.53
38	0.58	0.21	620	7.49
39	0.62	0.21	622	7.42
40	0.67	0.21	623	7.39
41	0.71	0.21	625	7.35
42	0.76	0.21	626	7.32
43	0.80	0.21	628	7.28
44	0.84	0.21	629	7.25
45	0.89	0.21	631	7.25
46	0.93	0.21	633	7.21
47	0.97	0.21	634	7.21
48	1.01	0.21	635	7.18
49	1.06	0.21	637	7.18
50	1.10	0.20	638	7.14
51	1.14	0.20	640	7.14
52	1.18	0.20	641	7.14
53	1.22	0.20	643	7.14
54	1.27	0.20	644	7.14
55	1.31	0.20	646	7.14
56	1.35	0.20	647	7.14
57	1.39	0.20	649	7.14
58	1.43	0.21	650	7.18
59	1.48	0.21	652	7.18
60	1.52	0.21	653	7.21
61	1.56	0.21	655	7.21
62	1.60	0.21	656	7.25
63	1.65	0.21	658	7.25
64	1.69	0.21	659	7.28
65	1.73	0.21	661	7.32
66	1.78	0.21	662	7.35
67	1.82	0.21	664	7.39
68	1.87	0.21	665	7.42
69	1.91	0.21	667	7.46
70	1.96	0.22	669	7.53
71	2.00	0.22	670	7.56
72	2.05	0.22	672	7.60
73	2.10	0.22	673	7.67
74	2.15	0.22	675	7.74
75	2.20	0.22	677	7.81
76	2.25	0.22	679	7.84
77	2.30	0.23	680	7.95
78	2.35	0.23	682	8.02
79	2.40	0.23	684	8.09
80	2.46	0.23	686	8.19
81	2.51	0.24	688	8.30
82	2.57	0.24	690	8.37
83	2.63	0.24	692	8.51
84	2.69	0.25	694	8.61
85	2.75	0.25	696	8.75
86	2.81	0.25	698	8.89
87	2.88	0.26	701	9.03
88	2.95	0.26	703	9.21

89	3.02	0.27	706	9.38
90	3.09	0.27	708	9.56
91	3.17	0.28	711	9.77
92	3.25	0.29	714	10.01
93	3.33	0.29	717	10.26
94	3.42	0.30	720	10.54
95	3.51	0.31	723	10.89
96	3.61	0.32	726	11.24
97	3.72	0.33	730	11.66
98	3.84	0.35	734	12.15
99	3.96	0.36	739	12.71
100	4.10	0.38	744	13.37
101	4.26	0.41	749	14.21
102	4.43	0.44	755	15.23
103	4.64	0.47	762	16.52
104	4.89	0.52	771	18.31
105	5.20	0.60	782	20.97
106	5.63	0.73	797	25.41
107	6.35	1.01	822	35.49
108	7.76	2.01	872	70.25

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	-4.19	2.02	453	70.77
1	-2.74	1.04	504	36.51
2	-1.96	0.77	532	26.78
3	-1.47	0.64	549	22.54
4	-1.10	0.57	562	20.06
5	-0.80	0.53	572	18.38
6	-0.54	0.49	581	17.15
7	-0.31	0.46	589	16.24
8	-0.11	0.44	596	15.51
9	0.08	0.43	603	14.95
10	0.26	0.42	609	14.56
11	0.43	0.41	615	14.25
12	0.59	0.40	621	14.07
13	0.75	0.40	626	13.97
14	0.91	0.40	632	13.93
15	1.07	0.40	638	13.93
16	1.23	0.40	643	14.00
17	1.39	0.40	649	14.11
18	1.56	0.41	655	14.21
19	1.73	0.41	660	14.39
20	1.90	0.42	666	14.56
21	2.07	0.42	673	14.81
22	2.26	0.43	679	15.12
23	2.45	0.44	686	15.51
24	2.65	0.46	693	16.00
25	2.87	0.48	700	16.66
26	3.11	0.50	709	17.54
27	3.37	0.53	718	18.69
28	3.68	0.58	729	20.37
29	4.06	0.65	742	22.89
30	4.57	0.78	760	27.20
31	5.37	1.06	788	37.00
32	6.85	2.03	840	71.12

APPENDIX C: STANDARD SETTING MATERIALS

C.1: Standard Setting Meeting Two-Day Agenda

Tuesday, June 6, 2006

- 8:00 – 8:30 a.m. *Continental Breakfast* — Pueblo/Sonora Room
- 8:30 – 9:45 a.m. General Session — Pueblo/Sonora Room
 Welcome and Introductions
 Standard Setting Training
- 9:45 – 10:00 a.m. *Break (Refreshments)* — Pueblo/Sonora Room
- 10:00 – 12:00 a.m. Individual Groups Review and Discuss Threshold Descriptors
 — Sedona 1, 2, 3, 4 and Phoenix 1 Rooms
- 12:00 – 1:00 p.m. *Lunch* — Kiva Room
- 1:00 – 3:00 p.m. Take Test — Sedona 1, 2, 3, 4 and Phoenix 1 Rooms
- 3:00 – 3:15 p.m. *Break (Refreshments)* — Pueblo/Sonora Room
- 3.15 – 5:00 p.m. Round 1 Item Ratings — Sedona 1, 2, 3, 4 and Phoenix 1 Rooms

Wednesday, June 7, 2006

- 8:00 – 8:30 a.m. *Continental Breakfast* — Pueblo/Sonora Room
- 8:30 – 11:00 a.m. Round 1 Results and Discussion and Round 2 Item Ratings
 — Sedona 1, 2, 3, 4 and Phoenix 1 Rooms
- 9:45 – 10:00 a.m. *Break (Refreshments)* — Pueblo/Sonora Room
- 11:00– 2:00 p.m. *Data Entry*
- 12:00 – 1:00 p.m. *Lunch* — Kiva Room
- 2:00 – 3:00 p.m. Round 2 Results and Discussion
 — Sedona 1, 2, 3, 4 and Phoenix 1 Rooms
- 3:00 – 3:15 p.m. *Break (Refreshments)* — Pueblo/Sonora Room
- 3:00 – 3:30 p.m. Round 3: Adjust Final Cut Points
 — Sedona 1, 2, 3, 4, and Phoenix 1 Rooms
- 3:30 – 3:45 p.m. Evaluation of Standard Setting
 — Sedona 1, 2, 3, 4, and Phoenix 1 Rooms

C.2: Performance Level Descriptors

Preliteracy (ELL I -- Kindergarten)

Performance Level Descriptions (in English)

Detailed below are the five performance level descriptions for each skill, which are based on the student's scaled score for each subtest.

LISTENING

Pre-Emergent. This student's Listening Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English.

Emergent. This student's Listening Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English.

Basic. This student's Listening Performance Level is Basic. This student understands a limited number of common words and simple phrases on topics of personal relevance when spoken slowly with extensive rephrasing, frequent repetitions, and contextual clues. This student is able to identify by name a few familiar objects, people, and events. This student comprehends and follows simple routine instructions for classroom activities that depend on gestures and other contextual clues.

Intermediate. This student's Listening Performance Level is Intermediate. This student understands a few common words and simple phrases on topics of personal relevance and may need frequent rephrasing, repetition, and contextual clues to increase comprehension. This student can identify by name some familiar objects, people, and events. This student comprehends and follows short routine instructions (2- to 5-word phrases) for classroom activities in the presence of gestures and clear contextual clues.

Proficient. This student's Listening Performance Level is Proficient. This student understands some words, phrases, and short sentences on topics of personal relevance when spoken slowly with some rephrasing, repetitions, and contextual clues. This student can identify by name many familiar objects, people, and events. This student is able to identify the initial and final sounds (not letters) of a spoken word. This student comprehends and follows routine (2- to 3-step) instructions for classroom activities in the presence of gestures and clear contextual cues.

SPEAKING

Pre-Emergent. This student's Speaking Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little or no ability to speak in English.

Emergent. This student's Speaking Performance Level is Emergent. This student may try to communicate with gestures and other nonverbal methods, or may use a language other than English. This student has very limited ability to speak in English.

Basic. This student's Speaking Performance Level is Basic. This student speaks in isolated words (usually a single noun or verb), depending heavily on gestures to express meaning. This student can identify by name a few familiar objects, people, and events. This student is

able to produce English graphemes that correspond to graphemes the student already hears and produces in his or her first language.

Intermediate. This student's Speaking Performance Level is Intermediate. This student speaks in isolated words or strings of 2 to 3 words and depends on gestures to express meaning. This student speaks using limited grammatical structure and linguistic forms. This student can identify by name some familiar objects, people, and events. This student is able to produce English graphemes that correspond to graphemes the student already hears and produces in his or her first language, including initial and final consonants.

Proficient. This student's Speaking Performance Level is Proficient. This student speaks in short patterns of words and phrases using grade-appropriate English grammatical structures and linguistic forms. This student can identify by name many familiar objects, people, and events. This student is able to produce some English graphemes that do not correspond to graphemes the student already hears and produces in his or her first language, including long and short vowels.

SOCIAL/ORAL (LISTENING & SPEAKING)

Pre-Emergent. This student's Social Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English. This student has very little or no ability to speak in English.

Emergent. This student's Social Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English. This student has very limited ability to speak in English.

Basic. This student's Social Performance Level is Basic. This student comprehends and responds to greetings and leave-taking with simple words, gestures, and other nonverbal behavior. This student speaks in isolated words (usually a single noun or verb), depending heavily on gestures to express meaning. This student is able to produce English graphemes that correspond to graphemes the student already hears and produces in his or her first language.

Intermediate. This student's Social Performance Level is Intermediate. This student uses common social greetings and simple repetitive phrases using isolated words or strings of 2- to 3-word responses. This student uses simple vocabulary needed to respond to greetings, courtesy, and leave-taking. This student is able to produce English graphemes that correspond to graphemes the student already hears and produces in his or her first language, including initial and final consonants.

Proficient. This student's Social Performance Level is Proficient. This student responds to greetings, courtesy, and leave-taking. This student speaks in short patterns of words and phrases using grade-appropriate English grammatical structures and linguistic forms. This student uses accurate, purposeful, yet restricted vocabulary needed to ask and answer basic questions about personal information, and give and follow simple directions and imperatives, including warnings. This student is able to produce some English graphemes that do not correspond to graphemes the student already hears and produces in his or her first language, including long and short vowels.

PREREADING

Pre-Emergent. This student's Reading Performance Level is Pre-Emergent. This student made very few or no responses. This student has little or no knowledge of written English.

Emergent. This student's Reading Performance Level is Emergent. This student may be able to understand visual universal symbols and graphics associated with a text. This student understands almost no written English or only a few isolated words.

Basic. This student's Reading Performance Level is Basic. This student can identify and sort a few common objects and pictures into basic categories. This student is able to identify a few common signs, symbols, labels, and captions in the environment, including traffic signs. This student comprehends and follows simple 1-word written directions for classroom activities that are accompanied by picture cues.

Intermediate. This student's Reading Performance Level is Intermediate. This student can identify and sort some common objects into basic categories. This student is able to identify some common signs, symbols, labels, and captions in the environment. This student comprehends and follows simple 1-step (2- to 3-words) written directions for classroom activities that are accompanied by picture cues.

Proficient. This student's Reading Performance Level is Proficient. This student can identify and sort many common objects into basic categories. This student is able to identify many common signs, symbols, labels, and captions in the environment. This student comprehends and follows simple 1- to 2-step (2- to 5-words) written directions for classroom activities that are accompanied by picture cues.

COMPREHENSION (LISTENING AND READING)

Pre-Emergent. This student's Comprehension Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English and has little or no knowledge of written English.

Emergent. This student's Comprehension Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Basic. This student's Comprehension Performance Level is Basic. This student understands a limited number of common words and simple phrases on topics of personal relevance when spoken slowly with extensive rephrasing, frequent repetitions, and contextual clues. This student comprehends and follows simple routine oral instructions for classroom activities that depend on gestures and other contextual clues and simple 1-word written directions that are accompanied by picture cues.

Intermediate. This student's Comprehension Performance Level is Intermediate. This student understands a few common words and simple phrases on topics of personal relevance and may need frequent rephrasing, repetition, and contextual clues to increase comprehension. This student comprehends and follows short routine oral instructions (2- to 5-word phrases) for classroom activities in the presence of gestures and clear contextual clues and simple 1-step (2- to 3-words) written directions for classroom activities that are accompanied by picture cues.

Proficient. This student's Comprehension Performance Level is Proficient. This student understands some words, phrases, and short sentences on topics of personal relevance when spoken slowly with some rephrasing, repetitions, and contextual clues. This student comprehends and follows routine (2- to 3-step) oral instructions for classroom activities in the presence of gestures and clear contextual clues and simple 1- to 2-step (2- to 5-words) written directions for classroom activities that are accompanied by picture cues.

TOTAL WRITING (WRITING CONVENTIONS AND WRITING)

Pre-Emergent. This student's Total Writing Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand English and cannot write in English.

Emergent. This student's Total Writing Performance Level is Emergent. This student has almost no knowledge of the English alphabet. This student is very little or no ability to write single letters or words in English.

Basic. This student's Total Writing Performance Level is Basic. This student responds with drawings to stories dramatized or contextualized. This student is able to write, with support, 1 to 5 letters of the alphabet. This student is able to write, with support, his or her first name.

Intermediate. This student's Total Writing Performance Level is Intermediate. This student relates short messages by drawing, by using imitative writing, or by writing key, self-selected words. This student is able to independently and legibly write 1 to 5 letters of the alphabet. This student can write, with support, 5 to 10 letters of the alphabet legibly. Occasionally this student is able to write letters of given sounds. This student can write his or her first name.

Proficient. This student's Total Writing Performance Level is Proficient. This student relates messages by drawing, by using imitative writing, by dictating to an adult, or by writing key, self-selected words. This student is able to independently and legibly write 6 to 8 letters of the alphabet. This student can write, with support, 11 to 16 letters of the alphabet legibly. Sometimes this student is able to write letters of given sounds and write self-selected key words. Sometimes this student organizes writing from left to right and top to bottom.

Primary (ELL II -- Grades 1–2)

Performance Level Descriptions (in English)

Detailed below are the five performance level descriptions for each skill, which are based on the students' scaled score for each subtest.

LISTENING

Pre-Emergent. This student's Listening Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English.

Emergent. This student's Listening Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English.

Basic. This student's Listening Performance Level is Basic. This student comprehends key words, phrases, and most short sentences in simple predictable conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student comprehends and follows 1-step directions related to the position of one's movement in space when accompanied by contextual cues and gestures. This student is able to answer basic questions about read-aloud stories with 1- or 2-word responses.

Intermediate. This student's Listening Performance Level is Intermediate. This student comprehends a range of expressions used to request personal details, direct classroom activities, identify people, objects, and events, ask for permission, and grant permission when spoken slowly with some repetitions and contextual clues. This student comprehends and follows 2- to 3-step directions related to the position of one's movement in space when accompanied by contextual cues and gestures. This student can identify specific details of text read to him or her.

Proficient. This student's Listening Performance Level is Proficient. This student comprehends and follows short predictable discourse on familiar matters, including familiar events, routines, objects, and people, and likes, dislikes, wants, and feelings when spoken slowly with some repetitions and contextual clues. This student is able to identify the main idea of expository or functional text read to him or her. This student comprehends and follows 3- to 4-step directions related to the position of one's movement in space. This student is able to respond to simple questions about text read to him or her.

SPEAKING

Pre-Emergent. This student's Speaking Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little or no ability to speak in English.

Emergent. This student's Speaking Performance Level is Emergent. This student may try to communicate with gestures and other nonverbal methods, or may use a language other than English. This student has very limited ability to speak in English.

Basic. This student's Speaking Performance Level is Basic. This student speaks using below-grade level English grammatical structures and linguistic forms. Errors and pronunciation difficulties still may impede communication. This student is able to describe a person or object in some detail. This student can answer basic questions about read-aloud stories with

1- or 2-word responses. This student is able to issue 2- to 3-word basic, routine directions and commands in a manner that the listener can follow, although meaning may be conveyed by gestures.

Intermediate. This student's Speaking Performance Level is Intermediate. This student speaks using below-grade level English grammatical structures and linguistic forms. Many errors or irregular forms often impede communication. This student is able to ask and answer questions about the size, color, shape, physical characteristics, and number of familiar objects, using accurate and somewhat limited vocabulary. This student is able to retell a simple story, placing events in sequence, using key words, phrases, and simple sentences. This student is able to issue single-step directions and commands in a manner that the listener can follow, with less reliance on gestures to convey meaning.

Proficient. This student's Speaking Performance Level is Proficient. This student speaks using grade-appropriate grammatical structures and linguistic forms; however, errors sometimes impede communication. This student is able to ask and answer questions about various attributes of people and objects, using purposeful and somewhat varied vocabulary. This student is able to relate simple stories using logical organization and some descriptive words. This student is able to respond to stories by answering questions about cause and effect and other relationships. This student is able to issue 1- to 2-step routine directions in a manner that the listener can follow.

SOCIAL/ORAL (LISTENING AND SPEAKING)

Pre-Emergent. This student's Social Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English. This student has very little or no ability to speak in English.

Emergent. This student's Listening Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English. This student has very limited ability to speak in English.

Basic. This student's Social Performance Level is Basic. This student comprehends key words, phrases, and most short sentences in simple predictable conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student responds appropriately to most common social interactions. This student uses accurate but limited vocabulary. This student can communicate personal and survival needs and indicate problems in communicating in a number of ways. This student participates in short, routine social conversations with individuals in which he or she exchanges personal information and discusses personal needs.

Intermediate. This student's Social Performance Level is Intermediate. This student comprehends a range of expressions used to request personal details when spoken slowly with some repetitions and contextual clues. This student uses accurate, but ordinary and limited vocabulary. This student participates in social conversations in which he or she exchanges personal information and discusses personal experiences and needs. This student is able to ask for and grant permission, express ability to do or not do something, and, give and follow 1- to 2-step commands.

Proficient. This student's Social Performance Level is Proficient. This student comprehends short predictable discourses on familiar matters when spoken slowly with some repetitions

and contextual clues. This student uses accurate, purposeful and somewhat varied vocabulary. This student interacts with adults and peers in formal and informal settings, using English in socially and culturally appropriate ways. This student is able to participate in social conversations in which he or she discusses personal likes, dislikes, wants, and feelings. This student is able to indicate comprehension of a given situation, describe familiar events; state similarities and differences in objects and people; and, give and follow multiple step directions.

READING

Pre-Emergent. This student's Reading Performance Level is Pre-Emergent. This student made very few or no responses. This student has little or no knowledge of written English.

Emergent. This student's Reading Performance Level is Emergent. This student may be able to understand visual universal symbols and graphics associated with a text. This student understands almost no written English or only a few isolated words.

Basic. This student's Reading Performance Level is Basic. This student is able to identify letters, words, and sentences, and distinguish initial, medial, and final sounds in single-syllable words. This student recognizes a few common high frequency sight words. This student comprehends, with the aid of picture cues, a few simple content-area words. This student understands a couple of key words that signal grade-specific mathematics operations. This student is able to indicate the meaning of common signs, graphics, and symbols. This student comprehends and follows 2- to 5-word written directions for classroom activities when accompanied by picture cues.

Intermediate. This student's Reading Performance Level is Intermediate. This student is able to recognize some common high frequency sight words. This student is able to indicate the meaning of specific signs (e.g., traffic, safety, warning signs). Occasionally, this student is able to identify the words that comprise compound words. This student comprehends with the aid of picture cues some simple content-area words. This student understands a few key words that signal grade-specific mathematics operations and is able to comprehend a few simple math word problems. This student comprehends and follows short 2- to 3-step written directions for classroom activities when accompanied by some picture cues.

Proficient. This student's Reading Performance Level is Proficient. This student is able to recognize many common high frequency sight words. This student can use knowledge of inflectional endings to identify base words. Sometimes, this student is able to identify the words that comprise compound words. This student comprehends with the aid of picture cues many simple content-area words and a few, more complex words. This student understands some key words that signal grade-specific mathematics operations and is able to comprehend some simple math word problems. This student comprehends and follows up to 5-step written directions for classroom activities when accompanied by a few picture cues.

COMPREHENSION (LISTENING AND READING)

Pre-Emergent. This student's Comprehension Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand spoken English and has little or no knowledge of written English.

Emergent. This student's Comprehension Performance Level is Emergent. This student has very little ability to understand spoken English and understands only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Basic. This student's Comprehension Performance Level is Basic. This student comprehends key words, phrases, and most short sentences in simple predictable conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student comprehends and follows 1-step oral directions related to the position of one's movement in space when accompanied by contextual cues and gestures. This student comprehends with the aid of picture cues a few simple content-area words and understands a couple of key words that signal grade-specific mathematics operations.

Intermediate. This student's Comprehension Performance Level is Intermediate. This student comprehends a range of expressions used to request personal details, direct classroom activities, identify people, objects, and events, and ask for and grant permission when spoken slowly with some repetitions and contextual clues. This student comprehends and follows 2- to 3-step oral directions related to the position of one's movement in space when accompanied by contextual cues and gestures. This student comprehends, with the aid of picture cues, some simple content-area words. This student understands a few key words that signal grade-specific mathematics operations and is able to comprehend a few simple math word problems. This student comprehends and follows short 2- to 3-step written directions for classroom activities when accompanied by some picture cues.

Proficient. This student's Comprehension Performance Level is Proficient. This student comprehends and follows short predictable discourse on familiar matters, including familiar events, routines, objects, people, and also likes, dislikes, wants, and feelings when spoken slowly with some repetitions and contextual clues. This student comprehends and follows 3- to 4-step oral directions related to the position of one's movement in space. This student comprehends with the aid of picture cues many simple content-area words and a few more complex words. This student understands some key words that signal grade-specific mathematics operations and is able to comprehend some simple math word problems. This student comprehends and follows up to 5-step written directions for classroom activities when accompanied by a few picture cues.

TOTAL WRITING (WRITING CONVENTIONS AND WRITING)

Pre-Emergent. This student's Total Writing Performance Level is Pre-Emergent. This student made very few or no responses. This student does not understand English and cannot write in English.

Emergent. This student's Total Writing Performance Level is Emergent. This student has almost no knowledge of the English alphabet or an understanding of the English writing conventions of usage, mechanics, and spelling. This student has very little or no ability to write in English.

Basic. This student's Total Writing Performance Level is Basic. This student is able to write 2- to 3-word phrases and simple sentences using key words that are posted and commonly used in the classroom. This student is able to produce independent writing that controls for directionality (left to right, top to bottom), is written legibly, and leaves spaces between words. This student is able to independently and legibly write many letters of the alphabet

(upper case and lower case). Occasionally, this student is able to accurately write names and numbers with support. In informal writing, the student uses phonetic spellings, with the beginning phoneme correctly represented most of the time.

Intermediate. This student's Total Writing Performance Level is Intermediate. This student can recognize the distinguishing features of a sentence. This student is able to write a few familiar 3- to 4-word phrases about an event or character from a contextualized story. This student can recognize the distinguishing features of a sentence. This student is able to produce independent writing that uses basic grade-appropriate English conventions with many errors that may confuse the reader. This student can independently write all upper and lower case letters, attending to form and spatial alignment. Sometimes, this student is able to write, with support, numbers, letters, words, short phrases, and sentences for personal use, or to complete short writing tasks. In informal writing, this student uses phonetic spellings, with the beginning and final phonemes correctly represented most of the time.

Proficient. This student's Total Writing Performance Level is Proficient. This student is able to write several 3- to 4-word phrases and simple sentences about a personal experience. This student is able to produce independent writing that uses basic grade-appropriate English conventions with some errors and difficulty in naturalness of expression. Often, this student is able to accurately write, with support, numbers, letters, words, short phrases, and sentences for personal use, or to complete short writing tasks. In informal writing, this student uses phonetic spellings, with consonants correctly represented most of the time. This student is able to report events sequentially using a topic sentence and a concluding statement.

Elementary (ELL III -- Grades 3–5)

Performance Level Descriptions (in English)

Detailed below are the five performance level descriptions for each skill, which are based on the students' scaled score for each subtest.

LISTENING

Pre-Emergent. This student's Listening Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Listening Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues.

Basic. This student's Listening Performance Level is Basic. This student is able to recognize topics in read-aloud stories when spoken slowly with repetitions. This student is able to comprehend and follow 3- to 4-step directions related to the position of one's movements in space. This student can comprehend a few content-area words, including grade-level math and science vocabulary.

Intermediate. This student's Listening Performance Level is Intermediate. This student is able to identify basic facts from read-aloud stories and content area presentations with the assistance of contextual support and repetitions. This student is able to comprehend and follow 3- to 4-step directions related to the position, frequency, and duration of one's movements in space. This student can comprehend some content-area words, including grade-level math and science vocabulary.

Proficient. This student's Listening Performance Level is Proficient. This student is able to identify the factual details, key words and expressions, and overall gist of read-aloud stories and content area presentations with the assistance of contextual support and repetitions. Sometimes, this student comprehends and follows multiple-step instructions (4 or more steps) for familiar processes or procedures. This student can comprehend many content-area words, including grade-level math and science vocabulary.

SPEAKING

Pre-Emergent. This student's Speaking Performance Level is Pre-Emergent. This student made very few or no responses. This student may try to communicate with gestures and other nonverbal methods or may use a language other than English. This student has very limited or no ability to speak in English.

Emergent. This student's Speaking Performance Level is Emergent. This student has limited ability to speak in English. This student is able to issue 2- to 3-word basic, routine directions and commands in a manner that the listener can follow, although meaning may be conveyed by gestures.

Basic. This student's Speaking Performance Level is Basic. This student is able to speak using below-grade English grammatical structures and linguistic forms; errors and pronunciation difficulties may impede communication. This student uses accurate but limited vocabulary. This student is able to ask and answer basic instructional questions on the content presented, using words and phrases. This student can relate stories or events about routine activities, using logical organization. This student is able to indicate comprehension of a given situation, and to give and follow multiple step directions and commands.

Intermediate. This student's Speaking Performance Level is Intermediate. This student is able to speak using grade-appropriate English grammatical structures and linguistic forms; however, many errors or irregular forms often impede communication. This student uses accurate, but ordinary and somewhat limited, vocabulary. This student is able to ask and respond to basic instructional questions on the content presented using phrases and simple sentences. This student can relate stories or events about personal experiences, using logical organization and some descriptive vocabulary. This student is able to describe events and to state similarities and differences in objects.

Proficient. This student's Speaking Performance Level is Proficient. This student is able to speak using grade-appropriate English grammatical structures and linguistic forms however, errors sometimes impede communication. This student uses accurate, purposeful, and somewhat varied vocabulary. This student is able to ask and respond to instructional questions on the content presented using phrases and sentences. This student can present coherent personal narratives about ideas, events, or activities of interest, using logical organization. This student is able to use phrases and simple sentences, showing some evidence of connected discourse. This student is able to describe past events and to state intentions.

SOCIAL/ORAL (LISTENING AND SPEAKING)

Pre-Emergent. This student's Social Performance Level is Pre-Emergent. This student made very few or no responses. This student has very limited or no ability to speak in English. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Social Performance Level is Emergent. This student has limited ability to speak in English. This student is unable to speak using English grammatical structures and linguistic forms; many errors and pronunciation difficulties impede communication.

Basic. This student's Social Performance Level is Basic. This student is able to speak using below-grade English grammatical structures and linguistic forms; errors and pronunciation difficulties may impede communication. This student uses accurate but limited vocabulary. This student is able to participate in social conversations on topics of personal relevance and familiar events. This student is able to give and receive invitations and apologies, and express ability and inability to do or not do something.

Intermediate. This student's Social Performance Level is Intermediate. This student is able to speak using grade-appropriate English grammatical structures and linguistic forms however, many errors or irregular forms often impede communication. This student uses accurate, ordinary, and somewhat limited vocabulary. This student is able to participate in social conversations on familiar topics of personal reference. This student is able to discuss personal experiences, agree and disagree with others, and express personal feelings.

Proficient. This student's Social Performance Level is Proficient. This student is able to speak using grade-appropriate English grammatical structures and linguistic forms however, errors sometimes impede communication. This student uses accurate, purposeful, and somewhat varied vocabulary. This student can participate in social conversations by asking and responding to questions, providing advice, giving suggestions, describing past events, and posing hypotheticals.

READING

Pre-Emergent. This student's Reading Performance Level is Pre-Emergent. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Reading Performance Level is Emergent. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend with the aid of picture cues a few simple content-area words. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Reading Performance Level is Basic. This student is able to recognize some common roots and affixes when attached to known vocabulary. This student can identify the basic sequence of events and make relevant predictions about stories. This student can identify main ideas and key details of text. This student is able to understand a few words that indicate mathematics operations. This student is able to comprehend some simple grade-level math word problems. This student comprehends and follows up to 5-step written directions for classroom activities.

Intermediate. This student's Reading Performance Level is Intermediate. This student can paraphrase main points of a story that includes a scenario. This student can distinguish cause from effect in text and can identify the main ideas, key words, and important details in short text on a familiar topic. This student is able to recognize the difference between figurative and literal language. This student is able to understand some words that indicate mathematics operations. Occasionally, this student is able to comprehend grade-level math word problems. This student comprehends and follows a short set of written instructions on routine procedures.

Proficient. This student's Reading Performance Level is Proficient. This student is able to identify the components and main problem or conflict of a plot and its resolution. This student can identify the main ideas, key words, and important details in text that requires some level of inference. This student is able to identify stated cause and effect relationships in text. This student is able to understand many words that indicate mathematics operations. Sometimes, this student comprehends grade-level math word problems. This student comprehends and follows a set of written multi-step instructions on routine procedures.

COMPREHENSION (LISTENING AND READING)

Pre-Emergent. This student's Comprehension Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Comprehension Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend a few simple content-area words with the aid of picture cues. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Comprehension Performance Level is Basic. This student is able to comprehend and follow 3- to 4-step oral directions related to the position of one's movements in space. This student can comprehend a few content-area words, including grade-level math and science vocabulary. This student is able to understand a few words that indicate mathematics operations. This student is able to comprehend some simple grade-level math word problems. This student comprehends and follows up to 5-step written directions for classroom activities.

Intermediate. This student's Comprehension Performance Level is Intermediate. This student is able to comprehend and follow 3- to 4-step oral directions related to the position, frequency, and duration of one's movements in space. This student can comprehend some content-area words, including grade-level math and science vocabulary. This student is able to understand some words that indicate mathematics operations. Occasionally, this student is able to comprehend grade-level math word problems. This student comprehends and follows a short set of written instructions on routine procedures.

Proficient. This student's Comprehension Performance Level is Proficient. Sometimes, this student comprehends and follows multiple-step oral instructions (4 or more steps) for familiar processes or procedures. This student can comprehend many content-area words, including grade-level math and science vocabulary. This student is able to understand many words that indicate mathematics operations. Sometimes, this student comprehends grade-level math word problems. This student comprehends and follows a set of written multi-step instructions on routine procedures.

TOTAL WRITING (WRITING CONVENTIONS AND WRITING)

Pre-Emergent. This student's Total Writing Performance Level is Pre-Emergent. This student made very few or no responses. This student has almost no knowledge or understanding of the English writing conventions of usage, mechanics, and spelling. This student has very little or no ability to write in English.

Emergent. This student's Total Writing Performance Level is Emergent. This student has limited ability to write in English. This student is able to write isolated words, 2- to 3-word phrases, and simple sentences using key words that are posted and commonly used in the classroom. Occasionally, this student is able to write, with support, time, addresses, names, and numbers.

Basic. This student's Total Writing Performance Level is Basic. This student is able to produce independent writing that demonstrates satisfactory control over rudimentary grammatical structures. This student is able to write short, single paragraph personal narratives or friendly letters about topics and ideas that are broad and simplistic. This student is able to write with a voice that reads more like a report, and uses word choices that are

nonspecific and limited. This student's writing has little variation in sentence types and marginally recognizable internal structures or organization.

Intermediate. This student's Total Writing Performance Level is Intermediate. This student is able to produce independent writing that is written legibly, uses punctuation, capitalization, simple verb tenses, and subject-verb agreement with many errors that often impede communication. This student is able to write personal narratives or letters on familiar topics up to 2 paragraphs that have main ideas, although not well defined ones. This student's writing uses a recognizable introduction and conclusion although ideas are not always sequenced. This student is able to write using a voice that is rather mechanical. This student uses word choices that are accurate yet lack variety.

Proficient. This student's Total Writing Performance Level is Proficient. This student is able to produce independent writing that is written legibly, uses punctuation, capitalization, simple verb tenses, and subject-verb agreement with some errors that occasionally impede communication. This student is able to write essays and formal communications of up to 2 paragraphs in various genres that have identifiable main ideas and that contain general and supporting details. This student uses repetitive sentence patterns and occasionally attempts to use more complex sentences. This student's writing has simple organization with some relationship among ideas. This student is able to use a voice that shows a developing awareness of the audience. This student uses ordinary, generic word choices.

Middle Grades (ELL IV -- Grades 6–8)

Performance Level Descriptions (in English)

Detailed below are the five performance level descriptions for each skill, which are based on the student's scaled score for each subtest.

LISTENING

Pre-Emergent. This student's Listening Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Listening Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues.

Basic. This student's Listening Performance Level is Basic. This student is able to restate the gist of an oral discourse on personal, social, or grade-level academic topics, although repetition, rephrasing, and contextual support are required. This student is able to comprehend a few content area words, including a few grade-level math and science vocabulary words.

Intermediate. This student's Listening Performance Level is Intermediate. This student is able to paraphrase main ideas and the most important details in an oral discourse on personal, social, or grade-level academic topics, although some repetition, rephrasing, and contextual support is required. This student is able to comprehend some content area words, including some grade-level math and science vocabulary.

Proficient. This student's Listening Performance Level is Proficient. This student is able to summarize main ideas and supporting details in an oral discourse on personal, social, or academic topics, with little repetition or rephrasing required. This student can comprehend many content area words, including many grade-level math and science vocabulary words.

SPEAKING

Pre-Emergent. This student's Speaking Performance Level is Pre-Emergent. This student made very few or no responses. This student may try to communicate with gestures and other nonverbal methods or may use a language other than English. This student has very limited or no ability to speak in English.

Emergent. This student's Speaking Performance Level is Emergent. This student has limited ability to speak in English. This student is able to issue 2- to 3-word basic, routine directions and commands in a manner that the listener can follow, although meaning may be conveyed by gestures.

Basic. This student's Speaking Performance Level is Basic. This student is able to speak using satisfactory control over below-grade English grammatical structures and linguistic forms. This student can present information in coherent, connected discourse using accurate, but limited, vocabulary. This student is able to describe past events.

Intermediate. This student's Speaking Performance Level is Intermediate. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, many errors or irregular forms often impede communication. This student is able to use accurate, but ordinary and somewhat limited, vocabulary. This student can describe situations and events.

Proficient. This student's Speaking Performance Level is Proficient. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, habitual errors sometimes impede communication. This student is able to use accurate, purposeful, and somewhat varied vocabulary. This student can summarize events.

SOCIAL/ORAL (LISTENING AND SPEAKING)

Pre-Emergent. This student's Social Performance Level is Pre-Emergent. This student made very few or no responses. This student has very limited or no ability to speak in English. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Social Performance Level is Emergent. This student has limited ability to speak in English. This student is unable to speak using English grammatical structures and linguistic forms; many errors and pronunciation difficulties impede communication.

Basic. This student's Social Performance Level is Basic. This student is able to speak using satisfactory control over below-grade English grammatical structures and linguistic forms. This student uses accurate but limited vocabulary. This student is able to participate in social conversations, responding to questions and describing past events. This student is able to restate the gist of an oral discourse on personal, social, or grade-level academic topics, although repetition, rephrasing, and contextual support are required.

Intermediate. This student's Social Performance Level is Intermediate. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, many errors or irregular forms often impede communication. This student is able to use accurate, but ordinary and somewhat limited, vocabulary. This student is able to participate in social conversations, responding to questions and describing past events. This student is able to paraphrase main ideas and important details in an oral discourse, although some repetition, rephrasing, and contextual support are required.

Proficient. This student's Social Performance Level is Proficient. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, habitual errors sometimes impede communication. This student is able to use accurate, purposeful, and somewhat varied vocabulary. This student is able to participate in social conversations, responding to questions, expressing feelings, and reporting on events. This student is able to summarize main ideas and supporting details in an oral discourse, with little repetition or rephrasing required.

READING

Pre-Emergent. This student's Reading Performance Level is Pre-Emergent. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Reading Performance Level is Emergent. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend a few simple content-area words with the aid of picture cues. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Reading Performance Level is Basic. This student is able to comprehend and follow the sequence of narration in popular newspaper and magazine articles and popular easy fiction. This student knows the meaning of a few multiple-meaning words that have a different meaning in mathematics. This student comprehends and follows a set of written multi-step instructions.

Intermediate. This student's Reading Performance Level is Intermediate. This student knows the meaning of some multiple-meaning words that have a different meaning in mathematics. Occasionally, this student can read and comprehend a few grade-level mathematics word problems. Occasionally, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams. This student comprehends and follows a set of written multi-step instructions.

Proficient. This student's Reading Performance Level is Proficient. This student knows the meaning of many multiple-meaning words that have a different meaning in mathematics. This student is able to summarize main ideas in text. Sometimes, this student can read and comprehend some grade-level mathematics word problems. Sometimes, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

COMPREHENSION (LISTENING AND READING)

Pre-Emergent. This student's Comprehension Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Comprehension Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend, with the aid of picture cues, a few simple content-area words. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Comprehension Performance Level is Basic. This student is able to comprehend a few content area words, including a few grade-level math and science vocabulary words. This student understands the meaning of a few multiple-meaning words that have a different meaning in mathematics. This student comprehends sets of oral instructions related to tasks on familiar processes or procedures. This student is able to comprehend and follow the sequence of narration in popular newspaper and magazine articles and popular easy fiction. This student comprehends and follows a set of written multi-step instructions.

Intermediate. This student's Comprehension Performance Level is Intermediate. This student is able to comprehend some content area words, including some grade-level math and

science vocabulary. Occasionally, this student can read and comprehend a few grade-level mathematics word problems. Occasionally, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams. This student comprehends and follows a set of written multi-step instructions.

Proficient. This student's Comprehension Performance Level is Proficient. This student can comprehend many content area words, including many grade-level math and science vocabulary words. This student understands the meaning of many multiple meaning words that have a different meaning in mathematics. Sometimes, this student can read and comprehend some grade-level mathematics word problems. Sometimes, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

TOTAL WRITING (WRITING CONVENTIONS AND WRITING)

Pre-Emergent. This student's Total Writing Performance Level is Pre-Emergent. This student made very few or no responses. This student has almost no knowledge or understanding of the English writing conventions of usage, mechanics, and spelling. This student has very little or no ability to write in English.

Emergent. This student's Total Writing Performance Level is Emergent. This student has limited ability to write in English. This student is able to write isolated words, 2- to 3-word phrases, and simple sentences using key words that are posted and commonly used in the classroom. Occasionally, this student is able to write, with support, time, addresses, names, and numbers.

Basic. This student's Total Writing Performance Level is Basic. This student is able to produce independent writing that demonstrates satisfactory control over below-grade English conventions. This student is able to create essays in various genres that include topics and ideas that are broad and simplistic. This student is able to write with marginally recognizable internal structures or organization. This student uses word choices that are nonspecific and limited so at times it is hard to understand what the writer is trying to say. This student's writing has little variation in sentence types and a significant number of awkward or rambling constructions.

Intermediate. This student's Total Writing Performance Level is Intermediate. This student is able to produce independent writing that uses on-grade English conventions, and has many errors that often impede communication. This student is able to create essays in various genres that include identifiable main ideas although not defined meaningfully. This student is able to write with a recognizable introduction and conclusion although ideas not always sequenced. This student uses word choices that are accurate yet lack variety. This student's writing demonstrates satisfactory control over simple sentence structures.

Proficient. This student's Total Writing Performance Level is Proficient. This student is able to produce independent writing that uses on-grade English conventions, and has some errors that occasionally impede communication. This student is able to create essays in various genres that include identifiable main ideas that contain general supporting details. This student is able to write essays that have simple organization, with some relationship among ideas present and lapses in sequencing and use of transitions. This student uses ordinary, generic word choices and repetitive sentence patterns. Occasionally, this student attempts to write more complex sentence structures.

High School (ELL V -- Grades 9–12)

Performance Level Descriptions (in English)

Detailed below are the five performance level descriptions for each skill, which are based on the student's scaled score for each subtest.

LISTENING

Pre-Emergent. This student's Listening Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Listening Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues.

Basic. This student's Listening Performance Level is Basic. This student is able to paraphrase main ideas and the most important details in an oral discourse on personal, social, or grade-level academic topics, although some repetition, rephrasing, and contextual support is required. This student is able to comprehend some content area words, including some grade-level math and science vocabulary.

Intermediate. This student's Listening Performance Level is Intermediate. This student is able to summarize main ideas and supporting details in an oral discourse on personal, social, or academic topics, with little repetition or rephrasing required. This student can comprehend many content area words, including many grade-level math and science vocabulary words.

Proficient. This student's Listening Performance Level is Proficient. This student is able to respond to requests for facts and evaluate opinions, attitudes, and point of view of speakers in a broad range of persuasive and expressive personal, social, and academic topics. This student can comprehend content area words, including grade-level math and science vocabulary.

SPEAKING

Pre-Emergent. This student's Speaking Performance Level is Pre-Emergent. This student made very few or no responses. This student may try to communicate with gestures and other nonverbal methods or may use a language other than English. This student has very limited or no ability to speak in English.

Emergent. This student's Speaking Performance Level is Emergent. This student has limited ability to speak in English. This student is able to issue 2- to 3-word basic, routine directions and commands in a manner that the listener can follow, although meaning may be conveyed by gestures.

Basic. This student's Speaking Performance Level is Basic. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, many errors or irregular forms often impede communication. This student is able to use accurate, but ordinary and somewhat limited, vocabulary. This student can describe situations and events.

Intermediate. This student's Speaking Performance Level is Intermediate. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, habitual errors sometimes impede communication. This student is able to use accurate, purposeful, and somewhat varied vocabulary. This student can summarize events.

Proficient. This student's Speaking Performance Level is Proficient. This student is able to speak, using on-grade English grammatical structures and linguistic forms however occasional errors occur. This student uses accurate, precise, descriptive, and extensive vocabulary. This student can summarize events and draw inferences. This student is able to restate newly-learned information.

SOCIAL/ORAL (LISTENING AND SPEAKING)

Pre-Emergent. This student's Social Performance Level is Pre-Emergent. This student made very few or no responses. This student has very limited or no ability to speak in English. This student may try to communicate with gestures or in a language other than English.

Emergent. This student's Social Performance Level is Emergent. This student has limited ability to speak in English. This student is unable to speak using English grammatical structures; many errors and pronunciation difficulties impede communication.

Basic. This student's Social Performance Level is Basic. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, many errors or irregular forms often impede communication. This student is able to use accurate, but ordinary and somewhat limited, vocabulary. This student is able to participate in social conversations, responding to questions and describing past events. This student is able to paraphrase main ideas and important details in an oral discourse, although some repetition, rephrasing, and contextual support are required.

Intermediate. This student's Social Performance Level is Intermediate. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however, habitual errors sometimes impede communication. This student is able to use accurate, purposeful, and somewhat varied vocabulary. This student is able to participate in social conversations, responding to questions, expressing feelings, and reporting on events. This student is able to summarize main ideas and supporting details in an oral discourse, with little repetition or rephrasing required.

Proficient. This student's Social Performance Level is Proficient. This student is able to speak, using on-grade English grammatical structures and linguistic forms; however occasional errors occur. This student uses accurate, precise, descriptive, and extensive vocabulary. This student is able to participate in social conversations by responding to questions and expressing feelings; summarize events; and, report on events. This student is able to respond to requests for facts.

READING

Pre-Emergent. This student's Reading Performance Level is Pre-Emergent. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Reading Performance Level is Emergent. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend with the aid of picture cues a few simple content-area words. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Reading Performance Level is Basic. This student knows the meaning of some multiple-meaning words that have a different meaning in mathematics. Occasionally, this student can read and comprehend a few grade-level mathematics word problems. Occasionally, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams. This student comprehends and follows a set of written multi-step instructions.

Intermediate. This student's Reading Performance Level is Intermediate. This student knows the meaning of many multiple-meaning words that have a different meaning in mathematics. This student is able to summarize main ideas in text. Sometimes, this student can read and comprehend some grade-level mathematics word problems. Sometimes, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

Proficient. This student's Reading Performance Level is Proficient. This student knows the meaning of most multiple-meaning words that have a different meaning in mathematics. This student is able to translate a written sentence or phrase into an algebraic equation or expression and can consistently comprehend most grade-level mathematics word problems. This student is able to consistently interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

COMPREHENSION (LISTENING AND READING)

Pre-Emergent. This student's Comprehension Performance Level is Pre-Emergent. This student made very few or no responses. This student has very little ability to understand spoken English and understands only a few isolated words. This student understands almost no written English or only a few isolated words. This student may be able to understand visual universal symbols and graphics associated with a text.

Emergent. This student's Comprehension Performance Level is Emergent. This student is able to comprehend a few key words, phrases, and short sentences in simple conversations on topics of immediate personal relevance when spoken slowly with frequent repetitions and contextual clues. This student is able to understand a few common high frequency sight words and simple sentences in English. This student is able to comprehend, with the aid of picture cues, a few simple content-area words. This student is able to indicate the meaning of some common signs, graphics, and symbols.

Basic. This student's Comprehension Performance Level is Basic. This student is able to comprehend some content area words, including some grade-level math and science vocabulary. Occasionally, this student can read and comprehend a few grade-level mathematics word problems. Occasionally, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

Intermediate. This student's Comprehension Performance Level is Intermediate. This student can comprehend many content area words, including many grade-level math and science vocabulary words. This student understands the meaning of many multiple meaning words that have a different meaning in mathematics. Sometimes, this student can read and

comprehend some grade-level mathematics word problems. Sometimes, this student is able to interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

Proficient. This student's Comprehension Performance Level is Proficient. This student can comprehend content area words, including grade-level math and science vocabulary. This student knows the meaning of most multiple-meaning words that have a different meaning in mathematics. This student is able to translate a written sentence or phrase into an algebraic equation or expression and can consistently read and comprehend most grade-level mathematics word problems. This student is able to consistently interpret graphic sources of information such as maps, charts, graphs, timelines, tables, and diagrams.

TOTAL WRITING (WRITING CONVENTIONS AND WRITING)

Pre-Emergent. This student's Total Writing Performance Level is Pre-Emergent. This student made very few or no responses. This student has almost no knowledge or understanding of the English writing conventions of usage, mechanics, and spelling. This student has very little or no ability to write in English.

Emergent. This student's Total Writing Performance Level is Emergent. This student has limited ability to write in English. This student is able to write isolated words, 2- to 3-word phrases, and simple sentences using key words that are posted and commonly used in the classroom. Occasionally, this student is able to write, with support, time, addresses, names, and numbers.

Basic. This student's Total Writing Performance Level is Basic. This student is able to produce independent writing that uses on-grade English conventions, and has many errors that often impede communication. This student is able to create essays in various genres that include identifiable main ideas although not defined meaningfully. This student is able to write essays that have recognizable introductions and conclusions, although ideas are not always sequenced. This student uses word choices that are accurate yet lack variety. This student's writing demonstrates satisfactory control over simple sentence structures.

Intermediate. This student's Total Writing Performance Level is Intermediate. This student is able to produce independent writing that uses on-grade English conventions, and has some errors that occasionally impede communication. This student is able to create essays in various genres that include identifiable main ideas that contain general supporting details. This student is able to write essays that have simple organization, with some relationship among ideas present and lapses in sequencing and use of transitions. This student uses ordinary, generic word choices and repetitive sentence patterns. Occasionally, this student attempts to write more complex sentence structures.

Proficient. This student's Total Writing Performance Level is Proficient. This student is able to produce independent writing that uses on-grade English conventions and has only minor errors that do not impede readability. This student is able to create essays in various genres that include clear and focused main ideas that contain relevant supporting details. This student is able to write essays that have an organization that enhances the central ideas and that have logical sequencing. This student uses varied, descriptive word choices that adequately convey meaning and uses a variety of sentence lengths, structures, and complexities.

C.3: AZELLA Form AZ-1 Standard Setting Summary Results in Raw Score Units

	No. of Committee Members	Emergent Cut Scores			Basic Cut Scores			Intermediate Cut Scores			Proficient Cut Scores		
		Round 1	Round 2	Round 3	Round 1	Round 2	Round 3	Round 1	Round 2	Round 3	Round 1	Round 2	Round 3
Kinder	15	Max = 72 points											
Range		2-18	1-13	1-11	7-28	5-22	5-22	19-43	18-41	18-42	25-69	25-62	29-65
Mean		5.9	5.8	5.3	15.1	14.2	14.5	29.7	28.3	28.3	44.5	43.5	42.8
SD		4.4	4.2	3.2	5.7	5.3	4.9	7.1	6.5	6.3	10.3	10.1	8.5
SE _{Mean}		1.1	1.1	0.8	1.5	1.4	1.3	1.8	1.7	1.6	2.7	2.6	2.2
Median		4	5	5	15	14	14	29	27	27	43	43	43
SE _{Median}		1.4	1.3	1.0	1.8	1.7	1.6	2.3	2.1	2.1	3.3	3.3	2.8
IQR*		3.5	6.0	4.0	4.5	7.0	3.5	7.5	8.0	6.5	10.0	7.5	7.0
2nd Grade	15	Max = 100 points											
Range		4-48	4-25	4-25	19-65	19-41	19-50	41-79	48-63	50-75	52-98	79-91	81-95
Mean		14.1	9.7	15.8	32.5	31.1	38.4	53.8	56.4	62.4	77.9	83.4	87.6
SD		11.7	6.4	6.5	11.2	6.7	7.2	9.9	4.0	6.6	11.3	4.1	4.1
SE _{Mean}		3.0	1.7	1.7	2.9	1.7	1.9	2.5	1.0	1.7	2.9	1.1	1.1
Median		8	7	15	32	32	40	54	56	60	79	83	85
SE _{Median}		3.8	2.1	2.1	3.6	2.2	2.3	3.2	1.3	2.1	3.6	1.3	1.3
IQR		12.0	9.8	8.8	13.0	10.5	5.0	11.0	5.0	8.0	8.0	6.3	5.0
4th Grade	18	Max = 100 points											
Range		7-29	6-27	17-25	19-50	32-46	35-46	39-74	54-77	56-75	69-92	76-92	76-90
Mean		18.6	20.4	21.2	36.2	40.6	41.6	58.2	64.2	66.1	79.7	84.6	86.1
SD		6.7	5.2	2.6	9.4	4.6	3.2	8.7	6.2	4.9	7.2	5.1	4.1
SE _{Mean}		1.6	1.2	0.6	2.2	1.1	0.8	2.0	1.5	1.1	1.7	1.2	1.0
Median		19	20	20	34	41	42	58	67	67	81	86	87
SE _{Median}		2.0	1.5	0.8	2.8	1.4	0.9	2.6	1.8	1.4	2.1	1.5	1.2
IQR		8.0	5.3	3.8	15.5	6.8	4.0	10.8	7.5	4.5	11.0	6.5	5.0
7th Grade	15	Max = 110 points											
Range		24-58	25-49	30-45	41-74	46-74	55-64	60-98	67-92	75-86	77-123	86-103	97-108
Mean		35.3	34.5	33.5	57.9	58.1	58.1	78.5	79.6	81.6	94.1	95.6	100.3
SD		11.3	7.8	5.0	9.7	7.2	2.9	9.1	6.8	4.1	10.7	5.9	2.4
SE _{Mean}		2.9	2.0	1.3	2.5	1.9	0.8	2.3	1.8	1.0	2.8	1.5	0.6
Median		30	32	30	58	57	57	76	81	83	94	97	100
SE _{Median}		3.7	2.5	1.6	3.1	2.3	0.9	2.9	2.2	1.3	3.5	1.9	0.8
IQR		18.0	12.3	6.5	12.0	9.0	4.8	9.0	9.5	5.0	7.5	10.0	0.0
9th Grade	14	Max = 110 points											
Range		26-65	27-73	30-40	51-80	52-88	60-70	74-93	76-97	80-97	86-104	96-105	100-105
Mean		40.0	38.4	36.8	62.9	63.4	62.3	84.0	85.3	85.4	100.1	101.4	102.1
SD		12.5	10.8	2.3	10.2	9.2	3.0	6.1	5.1	5.1	5.0	2.3	1.5
SE _{Mean}		3.3	2.9	0.6	2.7	2.5	0.8	1.6	1.4	1.4	1.3	0.6	0.4
Median		36	38	38	65	64	61	84	85	85	102	102	102
SE _{Median}		4.2	3.6	0.8	3.4	3.1	1.0	2.0	1.7	1.7	1.7	0.8	0.5
IQR		10.8	5.0	2.0	18.8	8.8	4.0	8.8	5.3	4.5	3.3	3.0	0.0

* IQR is Inter-quartile Range