ISBN 999-8516-33-1

9 789998 516335

1 2 3 4 5 6 7 8 9 10 11 12 A B C D E

Harcourt

AZELLA Technical Manual – Form AZ-1

Arizona

English Language Learner Assessment

Technical Manual

Harcourt

999-8516-33-1

About the Artist

NAME *Raelando Johnson*

AGE *13*

GRADE *Seventh*

LANGUAGE *Navajo*

TEACHER *Paul Williams, Jr.*

DISTRICT *Ganado Unified School District*

SCHOOL *Ganado Middle School*

CITY *Ganado*

Arizona

English Language Learner Assessment

Technical Manual

Harcourt

Table of Contents

Summary of Technical Report

This is a summary of the technical information for the AZELLA 2006 administration. It summarizes essential psychometric information including:

- classical item-level and subtests statistics;
- reliability;
- validity;
- calibration, equating, and scaling;
- IRT statistics;
- standard setting; and
- the summary of operational test results.

Classical Item-Level and Subtests Statistics

This section presents the item-level and subtests 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 (based on the Forms Field Test).

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, three indices were used:

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

Table 4.1 provides reliabilities by grade and by strand. The reliability of the Total Composite for grades K and 1 is 0.93, and the reliability for grades 3 through 12 is in the high 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 Decision at Proficient Cut

Based on the AZELLA 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.

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, 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, shown in Table 4.2.

Table 4.2 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. 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.

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 amount of students who achieved Proficient Status for the Total Test is less than 3% for grades K and 1. This is why the accuracy and consistency rates are 0.99 and 1.0, respectively.

Validity

For the 2006 administration of the AZELLA, Form A 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 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 supports the validity-related standards set forth in the *Standards for Educational and Psychological Testing*. 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)

  The alignment of the AZELLA Form AZ-1 to the AZ ELL standards was 85%.

- 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)

  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, except grade K.

  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" (p. 13)

Evidence of validity based on relationships to other variables will be conducted as ongoing research. Data from the 2007 administration of the AIMS will have to be obtained in order to examine the relationship.

Calibration, Equating, and Scaling

Harcourt Assessment 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, were fixed to the parameter values from the pre-existing vertical scale. Any remaining non-SELP items on the AZELLA were calibrated together with the SELP items using the Rasch and Partial Credit models. 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 ability metric.

Table 6.1 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 <0.5 and >1.5.

The OUTFIT and the INFIT statistics are presented in the item statistics tables in Appendix A. Using the cutoffs at <0.5 and >1.5, none of the items were flagged for INFIT, and the items flagged for OUTFIT (misfit) ranged from 1% to 11%.

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

Standard Setting

As the contractor for the AZELLA, Harcourt Assessment organized a performance standard setting meeting. Harcourt Assessment involved more than 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 8.2a and 8.2b show the scaled score cut 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:

Table 9.1 presents the percentages of students in each 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.

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 development, research, and scoring of the five grade spans and four subtests linked to the State's approved K–12 English language learner (ELL) proficiency standards. For the 2006 test administration, Harcourt Assessment, Inc., proposed the use of the Stanford English Language Proficiency Test (SELP), Form A, in conjunction with new items written by a team of Arizona educators under the facilitation of Aha!, Inc., a state contractor for facilitation, consulting, and technical writing services. The test was developed for five grade spans (K, 1–2, 3–5, 6–8, 9–12) and in four subtests (Speaking, Listening, Reading, and Writing) to assess the English language proficiency of students in kindergarten through grade 12 who are English language learners. The test is in accordance with the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999) and Arizona state testing requirements. The test is consistent with the principles of universal design and also consistent with applicable federal and state testing requirements.

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). 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' 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 the test are (1) to determine appropriate placement of students who have a Primary Home Language Other Than English (PHLOTE) and (2) to measure PHLOTE students' annual improvement in achieving English language proficiency. Thus, the test results provide the criteria for entry into Structured English Immersion (SEI) programs and the criteria for exiting SEI programs. English language learners deemed proficient through the test are reclassified as "fluent English proficient" and exit the SEI program. These students are then placed into fluent English proficient mainstream classrooms. Reclassified English language learners are tested annually at the end of each school year for two years. Students who fail to demonstrate proficiency will be reenrolled in SEI.

1.3 Recommended Test Use

The 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 Test Accommodations

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. Applying universal test design during the development process helped eliminate the need to address after-the-fact accommodations and provided a better assessment for all students. 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, in-house content and fairness experts reviewed the forms to ensure that concepts of universal design were incorporated. A stringent review of forms for special populations, such as the visually or hearing-impaired student, was conducted to ensure that items were fair, reliable, and accessible to all.

2. Test Design and Development

2.1 Test Specifications by Subtest by Grade Span

The 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 table 2.1, the number of items per grade span increases from K–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 band 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
- Prewriting items (kindergarten)
- Two extended responses to graphics-based prompts for grades 1 through 5 and one extended response to graphics-based prompts and one extended response without graphic-based prompts for grades 6 through 12

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

The test design for the 2006 administration of the AZELLA 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 A, 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 Subtests								
Grade Span	Speaking	Listening	Reading		Writing			Total Number of Items per Grade Span
					Writing Conventions	Prewriting	Writing Prompt	
	CR	MC	MC	Passages	MC	SR	ER	
K	12	12	12	0	0	16	1	53
1–2	14	20	20	9	20	0	2	76
3–5	14	20	20	4	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 Subtests								
Grade Span	Speaking	Listening	Reading		Writing			Total Number of Points per Grade Span
					Writing Conventions	Prewriting	Writing Prompt	
	CR	MC	MC	Passages	MC	SR	ER	
K	26	12	12	0	0	20	2(SR)	72
1–2	32	20	20	9	20	0	8	100
3–5	32	20	20	4	20	0	8	100
6–8	34	20	24	4	24	0	8	110
9–12	34	20	24	4	24	0	8	110

2.2 Item Development and Review Processes

In order to create a new and fully aligned assessment for English language learners for the 2006 administration, and also to meet the reporting requirements for NCLB in 2006, Harcourt Assessment made use of the SELP, Form A, and worked in conjunction with a team of Arizona educators under the facilitation of Aha!, Inc., to produce quality, customized items. The ADE reviewed the AZELLA forms prior to administration.

2.3 Scoring of the Writing Field Test

Field test student responses were reviewed, and three examples of appropriate anchor papers that represented each of the score points of 0 through 4 were selected. Additionally, several other excellent examples of each score point were presented to the state department in case substitutions for the original submissions were needed. After the anchors were selected, the remaining responses were used for Training Sets and Qualifying Sets.

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 Training Supervisor. This process was repeated for Training Set B. The Rater was then ready to qualify to score. In order to do so, raters must pass one Qualifying Set containing 10 papers with a score of 80% correct. Raters must take at least two Qualifying Sets even if they receive passing scores on both.

During the scoring of live student responses, experienced leaders check-score 10% of each rater's work product. The agreement rate between the two Raters' scores is tracked by computer reports. Feedback about accuracy and productivity is given to the raters every morning by the Project Training Supervisor based on the computer reports. Additionally, Raters can ask scoring questions of the Training Supervisor either electronically or in person.

2.4 Test Construction

For a week in August 2005, a group of Arizona educators met to augment Form A of the SELP. At least 30% of the Form A SELP questions were retained in each domain (Reading, Writing and Writing Conventions, Listening, and Speaking). The educators reviewed the retained items and developed new test items that were grade-band appropriate. They used the 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 convened by the ADE in September. Several items were modified, deleted, or replaced before all the test items were submitted to Harcourt Assessment for 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 field test was co-developed by the ADE and Harcourt Assessment.

The 2006 AZELLA 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 the AZELLA is the language that is being tested. 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 the AZELLA allows the test 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, the AZELLA also provides evidence of students' progress toward proficiency required by NCLB.

To determine the appropriate language for English language learner items and stimuli, Harcourt 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. Harcourt Assessment 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. Harcourt 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 have been the subject of special investigation at Harcourt Assessment. Harcourt Assessment's English language proficiency professionals individually administered a pilot of the Listening and Speaking tests to English language learners during cognitive labs, carefully observing and recording student responses and eliciting their reactions to the tests. Outcomes of the cognitive labs led to important design decisions regarding:

- item types,
- number of items,
- length of pauses between items,
- use of recorded stimuli, and
- recording student spoken responses.

The Listening and Speaking subtests of the 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 the 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 the AZELLA.

3. Item-Level Statistics

3.1 Classical Test Theory

There are useful indices available within the framework of classical test theory (CTT) for estimating the precision of the raw test scores and the reliability of assessments. Within CTT, an observed test score is defined as an imprecise estimate of a student's true (and unobservable) ability level and is composed of two components. The first component is referred to as "true score" and is the portion of the observed score that is directly dependent on the student's ability level. The second is an error component (error) and is the portion of the score that is attributable to random error; that is, the portion of the score attributable to factors unrelated to the student's ability. Error for any student is normally distributed around that student's true score with a mean of zero and an arbitrary standard deviation. Suppose it were possible to give an exam to one student a large number of times without any practice effects. If the resulting distribution of scores was to be examined, a normal distribution with a certain mean and a certain standard deviation about the mean would be found. The mean of the resulting distribution is the student's true score according to the definition of error given above. For each student who responds to the exam, error is normally distributed with a mean of zero. However, the standard deviation of the error distribution is idiosyncratic to each student (though it tends to be larger toward the low and high ends of the exam for most tests). If one wanted to estimate what would likely be the standard deviation of this distribution of errors for any arbitrary examinee, the best estimate would be the mean of the standard deviations of the error distribution across all examinees. This quantity is called the standard error of measurement (SEM), and is denoted as σ_E. It is defined as:

$$\sigma_E = \sigma_t \sqrt{1 - \rho_t}$$

where σ_t is the standard deviation of the raw scores for the exam and ρ_t is the reliability coefficient for the exam.

3.2 Item-Level Descriptive Statistics

This section presents the raw score summary statistics for all items in the AZELLA that were tested between February 20 and March 10, 2006 (Forms Field Testing). 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 does an item with 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, and
- Infit and Outfit statistics.

3.3 Subtest Statistics

The AZELLA scores are reported on the following strands: Listening, Speaking, Comprehension, Oral, Reading, Total Writing, and Total Test. 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

Level	Strand	Max Point Possible	Max Point Received	N-Count	Mean	SD	SE of Mean
Preliteracy	Listening	12	12	336	9.68	2.15	0.12
K	Reading	12	12	327	8.62	2.76	0.15
	Speaking	26	26	328	18.10	5.78	0.32
	Comprehension	24	24	327	18.33	4.25	0.23
	Oral	38	38	328	27.80	7.32	0.40
	Total Test	72	70	317	48.81	11.22	0.63
	Total Writing	22	21	333	12.08	3.70	0.20
Primary	Listening	20	20	565	17.01	2.35	0.10
1–2	Reading	20	20	571	12.04	4.47	0.19
	Speaking	32	32	565	24.98	7.43	0.31
	Comprehension	40	40	561	29.10	5.73	0.24
	Oral	52	52	557	42.00	8.95	0.38
	Total Test	100	98	554	68.90	15.64	0.66
	Total Writing	28	27	574	14.62	5.51	0.23
Elementary	Listening	20	20	489	13.98	3.92	0.18
3–5	Reading	20	20	484	11.40	4.56	0.21
	Speaking	32	32	470	23.61	8.18	0.38
	Comprehension	40	39	484	25.40	7.76	0.35
	Oral	52	51	468	37.64	11.06	0.51
	Total Test	100	96	463	68.26	19.32	0.90
	Total Writing	28	28	483	18.92	6.27	0.29
Middle Grades	Listening	20	20	461	13.11	3.93	0.18
6–8	Reading	24	24	462	13.36	5.01	0.23
	Speaking	34	34	459	24.92	9.31	0.43
	Comprehension	44	43	459	26.49	8.22	0.38
	Oral	54	54	454	38.00	12.37	0.58
	Total Test	110	106	452	72.20	21.37	1.01
	Total Writing	32	32	463	20.69	6.34	0.29
High School	Listening	20	20	1093	11.02	3.62	0.11
9–12	Reading	24	24	1088	12.58	5.13	0.16
	Speaking	34	34	1050	22.28	9.35	0.29
	Comprehension	44	41	1080	23.60	7.98	0.24
	Oral	54	53	1041	33.28	11.76	0.36
	Total Test	110	102	1035	63.81	20.49	0.64
	Total Writing	32	31	1091	17.81	6.02	0.18

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. Harcourt Assessment 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:

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

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, but 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 individual's observed total score is considered the estimate of the person's true score. Because the standard error of measurement is inversely related to the reliability of a test, the greater the reliability, the less the standard error of measurement and the more confidence one may have in the accuracy, or precision, of the 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:

$$SEM = SD\sqrt{1 - r_{xx}} \quad \Leftrightarrow \quad s_e = s_x\sqrt{1 - \frac{s_t^2}{s_x^2}}$$

where

SEM (= s_e) refers to the standard error of measurement,

SD (= s_x) is the standard deviation unit of the scale for a test,

r_{xx} is the reliability coefficient for a sample test (or estimate of ρ_{XX}, which is a population reliability coefficient),
s_t^2 is the estimate of σ_T^2, and
s_x^2 is the estimate of σ_X^2.

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 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.

A confidence band can be used in interpreting the ability estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by

$$\hat{\beta} \pm SEM .$$

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 Strands

Table 4.1 provides the raw score and scaled score descriptive statistics and reliabilities by grade by the reporting strands using the Wave 1 Operational 2006 data. It includes the following information:

- 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
- conditional standard errors of measurement, IRT based, for the Proficient cut scores only.

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

Kindergarten

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	12	12	28482	7.5	3.0	518.6	61.5	0.78	3.53	0.63	24.82
Prereading	12	12	28482	4.1	2.9	481.3	60.0	0.76	1.99	0.34	25.24
Speaking	12	26	28482	14.0	7.2	530.9	69.4	0.92	4.00	0.54	16.52
Comprehension	24	24	28482	11.6	5.1	500.2	52.3	0.84	4.70	0.48	18.94
Oral	24	38	28482	21.5	9.5	527.5	60.0	0.92	6.18	0.57	14.35
Prewriting	17	22	28482	5.8	4.3	465.2	73.7	0.87	2.08	0.26	23.24
Total Composite	53	72	28482	31.4	13.8	506.4	48.6	0.93	8.36	0.44	12.25

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

1st Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	8987	15.1	3.6	577.4	55.7	0.82	6.38	0.75	24.22
Reading	20	20	8987	7.1	3.9	555.0	51.4	0.76	3.46	0.35	20.20
Speaking	14	32	8987	19.3	9.7	578.1	75.1	0.95	4.28	0.60	15.33
Comprehension	40	40	8987	22.1	6.3	564.0	46.6	0.84	8.75	0.55	15.30
Oral	34	52	8987	34.3	12.1	578.0	57.2	0.93	8.89	0.66	12.57
Total Writing	22	28	8987	8.3	4.8	561.1	46.8	0.77	3.96	0.29	16.91
Total Composite	76	100	8987	49.7	17.4	568.5	46.0	0.93	12.91	0.50	9.87

2nd Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	7013	16.6	3.5	607.2	61.0	0.86	6.26	0.83	26.18
Reading	20	20	7013	11.9	4.9	607.9	60.9	0.86	4.40	0.59	22.44
Speaking	14	32	7013	23.3	9.2	607.9	78.7	0.95	4.97	0.73	16.59
Comprehension	40	40	7013	28.5	7.5	604.5	57.4	0.90	8.85	0.71	16.70
Oral	34	52	7013	39.8	11.8	604.7	61.9	0.94	9.64	0.77	13.79
Total Writing	22	28	7013	14.8	6.1	608.2	48.3	0.84	5.86	0.53	17.99
Total Composite	76	100	7013	66.5	20.1	603.3	53.6	0.95	14.19	0.67	10.50

3rd Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	5302	11.5	4.5	615.0	54.1	0.83	4.79	0.57	20.44
Reading	20	20	5302	9.2	4.6	615.6	57.3	0.84	3.72	0.46	20.62
Speaking	14	32	5302	22.1	9.9	637.6	89.4	0.96	4.41	0.69	15.40
Comprehension	40	40	5302	20.7	8.5	615.0	53.1	0.90	6.55	0.52	14.67
Oral	34	52	5302	33.6	13.3	624.6	61.4	0.94	8.05	0.65	12.39
Total Writing	22	28	5302	15.3	6.8	614.9	60.7	0.90	4.82	0.55	18.48
Total Composite	76	100	5302	58.1	22.7	618.5	57.0	0.96	11.35	0.58	9.56

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

4th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	4939	13.3	4.7	635.0	58.3	0.86	5.04	0.66	20.44
Reading	20	20	4939	10.9	5.0	634.4	61.6	0.87	3.99	0.54	20.62
Speaking	14	32	4939	23.1	10.0	647.5	92.8	0.96	4.32	0.72	16.42
Comprehension	40	40	4939	24.1	9.1	633.8	56.3	0.92	6.91	0.60	14.67
Oral	34	52	4939	36.3	13.7	638.6	64.8	0.95	8.13	0.70	12.85
Total Writing	22	28	4939	17.7	7.0	635.8	64.9	0.91	5.37	0.63	18.48
Total Composite	76	100	4939	64.9	23.6	634.2	58.8	0.97	11.81	0.65	9.91

5th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	4592	14.1	4.9	645.4	64.2	0.88	4.85	0.70	20.44
Reading	20	20	4592	12.1	5.3	648.2	68.0	0.89	4.03	0.60	20.62
Speaking	14	32	4592	23.7	10.1	654.5	95.8	0.97	4.24	0.74	16.42
Comprehension	40	40	4592	26.2	9.6	645.4	63.3	0.93	6.74	0.65	15.37
Oral	34	52	4592	37.8	14.1	647.1	70.7	0.96	7.85	0.73	13.41
Total Writing	22	28	4592	19.1	7.2	650.4	72.0	0.91	5.58	0.68	19.81
Total Composite	76	100	4592	68.9	24.9	644.7	66.0	0.97	11.52	0.69	10.36

6th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	4089	12.5	4.5	643.8	57.5	0.84	5.00	0.63	20.76
Reading	24	24	4089	13.5	6.0	649.8	61.3	0.89	4.46	0.56	18.45
Speaking	14	34	4089	25.5	10.4	670.2	87.2	0.97	4.73	0.75	14.88
Comprehension	44	44	4089	26.0	9.9	646.2	57.1	0.92	7.14	0.59	13.83
Oral	34	54	4089	38.1	13.9	653.1	62.4	0.95	8.58	0.71	12.32
Total Writing	26	32	4089	20.7	7.8	649.9	62.7	0.91	6.33	0.65	16.87
Total Composite	84	110	4089	72.2	25.6	649.4	58.4	0.97	12.89	0.66	8.86

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

7th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	4190	13.0	4.9	650.1	63.9	0.87	4.66	0.65	20.76
Reading	24	24	4190	14.2	6.3	656.3	67.0	0.91	4.32	0.59	18.45
Speaking	14	34	4190	24.6	11.1	664.3	92.8	0.97	4.26	0.72	15.86
Comprehension	44	44	4190	27.2	10.6	652.2	62.9	0.94	6.73	0.62	14.39
Oral	34	54	4190	37.7	15.0	653.0	69.0	0.96	7.92	0.70	12.78
Total Writing	26	32	4190	21.4	8.2	655.5	67.6	0.91	6.25	0.67	16.87
Total Composite	84	110	4190	73.2	27.7	651.9	64.8	0.97	12.07	0.67	9.35

8th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	3815	13.8	4.6	660.4	60.5	0.86	5.15	0.69	22.96
Reading	24	24	3815	15.3	6.1	668.1	64.6	0.91	4.72	0.64	19.95
Speaking	14	34	3815	25.3	10.7	671.2	89.9	0.97	4.55	0.75	17.15
Comprehension	44	44	3815	29.1	10.2	663.3	59.4	0.94	7.40	0.66	15.09
Oral	34	54	3815	39.1	14.1	661.5	63.7	0.95	8.67	0.72	13.93
Total Writing	26	32	3815	22.6	7.5	665.4	61.8	0.90	7.09	0.71	18.06
Total Composite	84	110	3815	77.1	25.7	661.4	57.5	0.97	13.49	0.70	9.98

9th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	5179	12.1	4.2	670.2	48.0	0.80	5.43	0.61	20.65
Reading	24	24	5179	14.1	5.7	675.3	55.4	0.88	4.93	0.59	18.31
Speaking	14	34	5179	24.8	10.9	687.9	91.7	0.97	4.38	0.73	14.60
Comprehension	44	44	5179	26.3	9.3	672.2	48.4	0.91	7.79	0.60	13.69
Oral	34	54	5179	36.9	14.1	674.9	57.4	0.95	8.56	0.68	12.22
Total Writing	26	32	5179	20.7	7.3	680.5	56.6	0.90	6.52	0.65	16.28
Total Composite	84	110	5179	71.7	25.3	675.0	52.4	0.97	13.05	0.65	8.86

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

10th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	2849	12.0	4.3	669.4	49.1	0.81	5.27	0.60	20.65
Reading	24	24	2849	14.5	5.7	679.2	55.5	0.88	5.02	0.60	18.31
Speaking	14	34	2849	24.2	11.4	685.4	95.5	0.97	4.11	0.71	14.60
Comprehension	44	44	2849	26.5	9.4	674.1	47.7	0.91	7.79	0.60	14.28
Oral	34	54	2849	36.2	14.5	673.1	58.4	0.95	8.37	0.67	12.67
Total Writing	26	32	2849	20.5	7.3	679.6	56.1	0.90	6.63	0.64	17.05
Total Composite	84	110	2849	71.2	25.6	674.8	52.0	0.97	13.02	0.65	9.14

11th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	1962	13.0	3.9	680.2	43.9	0.78	6.11	0.65	20.65
Reading	24	24	1962	16.0	5.3	694.1	52.1	0.87	5.85	0.67	19.85
Speaking	14	34	1962	26.3	9.8	705.0	85.2	0.96	5.06	0.77	15.51
Comprehension	44	44	1962	29.0	8.6	686.8	43.1	0.90	9.10	0.66	14.28
Oral	34	54	1962	39.3	12.5	686.5	50.8	0.94	9.91	0.73	13.20
Total Writing	26	32	1962	22.1	6.3	692.3	48.3	0.87	7.84	0.69	17.05
Total Composite	84	110	1962	77.4	22.2	688.0	44.4	0.96	15.51	0.70	9.49

12th Grade

Subtest	# of Items	Max Point	N-count	RS Mean	RS SD	SS Mean	SS SD	Reliability	SEM	Raw Score Proportion	Proficient Conditional SEM
Listening	20	20	1768	12.9	3.9	679.3	44.6	0.78	6.04	0.64	22.86
Reading	24	24	1768	16.1	5.3	694.6	51.1	0.87	5.86	0.67	19.85
Speaking	14	34	1768	26.3	9.6	704.3	82.4	0.96	5.36	0.77	16.73
Comprehension	44	44	1768	29.0	8.6	686.4	43.2	0.90	9.10	0.66	14.98
Oral	34	54	1768	39.2	12.3	686.1	49.9	0.93	10.09	0.73	13.83
Total Writing	26	32	1768	22.1	6.5	692.0	49.4	0.88	7.67	0.69	18.10
Total Composite	84	110	1768	77.4	22.1	687.7	43.8	0.96	15.66	0.70	9.91

4.5 Reliability of Classification Decision at Proficient Cut

Based on the AZELLA 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 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 second 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 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.

Table 4.2 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 Wave 1 AZELLA 2006 administration. The table contains the following:

- accurate classifications,
- false positives,
- false negatives, and
- 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 amount of students who achieved Proficient Status for the Total Test is less than 3% for grades K and 1. This is why the accuracy and consistency rates are 0.99 and 1.0, respectively.

Table 4.2: Decision and Consistency Table by Grade

Grade	Accuracy	False Positives	False Negatives	Consistency
K	0.99	0.01	0.00	0.98
1	1.00	0.00	0.00	1.00
2	0.90	0.11	0.00	0.86
3	0.88	0.07	0.05	0.85
4	0.87	0.05	0.09	0.81
5	0.89	0.04	0.07	0.84
6	0.91	0.04	0.05	0.87
7	0.92	0.03	0.05	0.88
8	0.88	0.05	0.07	0.83
9	0.92	0.04	0.05	0.88
10	0.91	0.04	0.05	0.88
11	0.91	0.04	0.04	0.88
12	0.89	0.04	0.07	0.85

5. Validity

For the 2006 administration of the AZELLA, Form A of the 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 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 supports the validity-related standards set forth in the *Standards for Educational and Psychological Testing*. 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), and
- 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, 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. Only those items relating specifically to an instructional standard were included in the test forms.

5.2 Evidence of the Test Content for AZELLA

In order for the 2006 AZELLA to accurately measure the English proficiency level of Arizona K–12 English language learners based on the Arizona English language learner (ELL) proficiency standards, the items in the Harcourt ELL item bank were reviewed to determine their match to individual performance indicators of Arizona ELL proficiency standards for each grade span. In addition, new items were created to better align the test to the standards. The alignment of the AZELLA Form AZ-1 to the Arizona ELL proficiency standards is 85%. This alignment study was conducted by an independent third party.

5.3 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.

5.4 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. 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, except for grade K. Students in Kindergarten do not usually read or write yet, but they can have Listening and Speaking skills.

The evidence of internal structure of the 2006 AZELLA is also depicted by the point biserial correlation coefficient and fit statistics. Appendices A.1–A.5 provide these statistics for the 2006 AZELLA.

Table 5.1: Intercorrelations Among the Subtests by Grade

Kindergarten

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.48	0.66	0.43
Reading		1.00	0.39	0.47
Speaking			1.00	0.40
Total Writing				1.00

Table 5.1: Intercorrelations Among the Subtests by Grade *(continued)*

1st Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.42	0.57	0.45
Reading		1.00	0.34	0.70
Speaking			1.00	0.41
Total Writing				1.00

2nd Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.55	0.64	0.60
Reading		1.00	0.51	0.81
Speaking			1.00	0.58
Total Writing				1.00

3rd Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.71	0.65	0.76
Reading		1.00	0.55	0.79
Speaking			1.00	0.65
Total Writing				1.00

4th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.74	0.68	0.79
Reading		1.00	0.57	0.80
Speaking			1.00	0.68
Total Writing				1.00

5th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.77	0.74	0.83
Reading		1.00	0.64	0.83
Speaking			1.00	0.75
Total Writing				1.00

6th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.76	0.70	0.80
Reading		1.00	0.60	0.83
Speaking			1.00	0.70
Total Writing				1.00

Table 5.1: Intercorrelations Among the Subtests by Grade *(continued)*

7th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.79	0.72	0.84
Reading		1.00	0.65	0.86
Speaking			1.00	0.74
Total Writing				1.00

8th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.78	0.66	0.82
Reading		1.00	0.61	0.84
Speaking			1.00	0.68
Total Writing				1.00

9th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.76	0.67	0.78
Reading		1.00	0.64	0.83
Speaking			1.00	0.76
Total Writing				1.00

10th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.75	0.63	0.79
Reading		1.00	0.63	0.85
Speaking			1.00	0.74
Total Writing				1.00

11th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.73	0.57	0.75
Reading		1.00	0.58	0.81
Speaking			1.00	0.69
Total Writing				1.00

12th Grade

	Listening	Reading	Speaking	Total Writing
Listening	1.00	0.72	0.57	0.74
Reading		1.00	0.58	0.82
Speaking			1.00	0.65
Total Writing				1.00

5.5 Evidence of 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) defines 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 factor to evaluate unidimensionality. If the ratio is large (usually larger than 3), the first factor is relatively strong.

The Wave 1 AZELLA 2006 administration (Operational) dataset was used to conduct the analyses. The file was cleaned by deleting students who did not take all four subtests. The analyses were carried out by test levels—Preliteracy, Primary, Elementary, Middle Grades, and High School. The Principal Components Analysis (PCA) method was used to extract total variance, and the results are shown in Table 5.2. The results in Table 5.2 support the following statements:

- There is one dominant factor in the AZELLA.
- The unidimensionality assumption for using the Rasch model is valid.

Table 5.2: Results of Factor Analysis

LEVEL	Eigen1	Eigen2	Eigen3	Ratio
Preliteracy	11.33	4.10	2.91	6.06
Primary	16.54	5.48	2.88	4.24
Elementary	21.73	4.41	2.01	7.19
Middle Grades	24.69	4.66	1.98	7.48
High School	23.08	4.01	1.84	8.79

5.6 Relationships to Other Variables

For the AZELLA, evidence of validity based on relationships to other variables will be conducted as ongoing research. Data from the 2007 administration of the AIMS will have to be obtained in order to examine the relationship of the AZELLA 2006 to the AIMS.

6. Calibration, Equating, and Scaling

The items on the AZELLA were analyzed within the framework of Item Response Theory (IRT). IRT is widely used because of the advantages it confers upon the exam consumers. It promotes equity of results from year to year through what has been referred to as test-free measurement. Simply stated, test-free measurement means that given a student's responses to two exams scaled using IRT, that student will achieve the same scaled score on both exams except for measurement error. This holds true regardless of differences in the overall difficulties of the exams. In other words, measurement is test-free in the sense that the results are dependent only upon the ability of the student and are independent of the item difficulties.

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. 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 was *Winsteps* Version 3.27 (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" while the curve on the right (j=1) shows the probability of getting a score of "1." 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%.

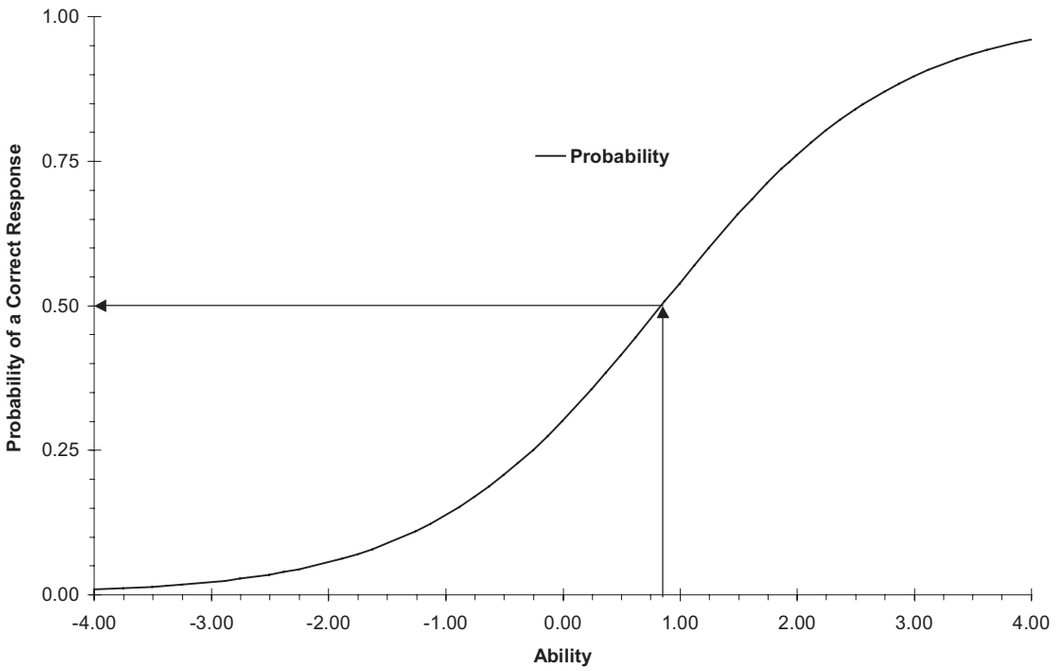

Figure 6.1: Sample Item Characteristic Curve

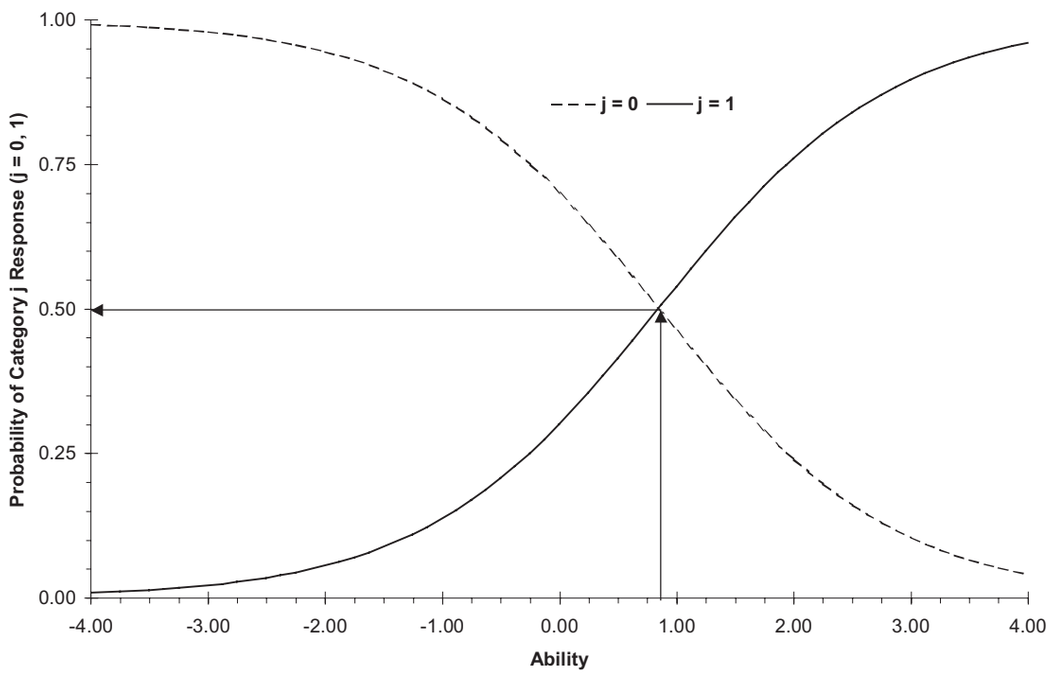

Figure 6.2: Category Response Curves for a One-step Item

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 PCM 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.

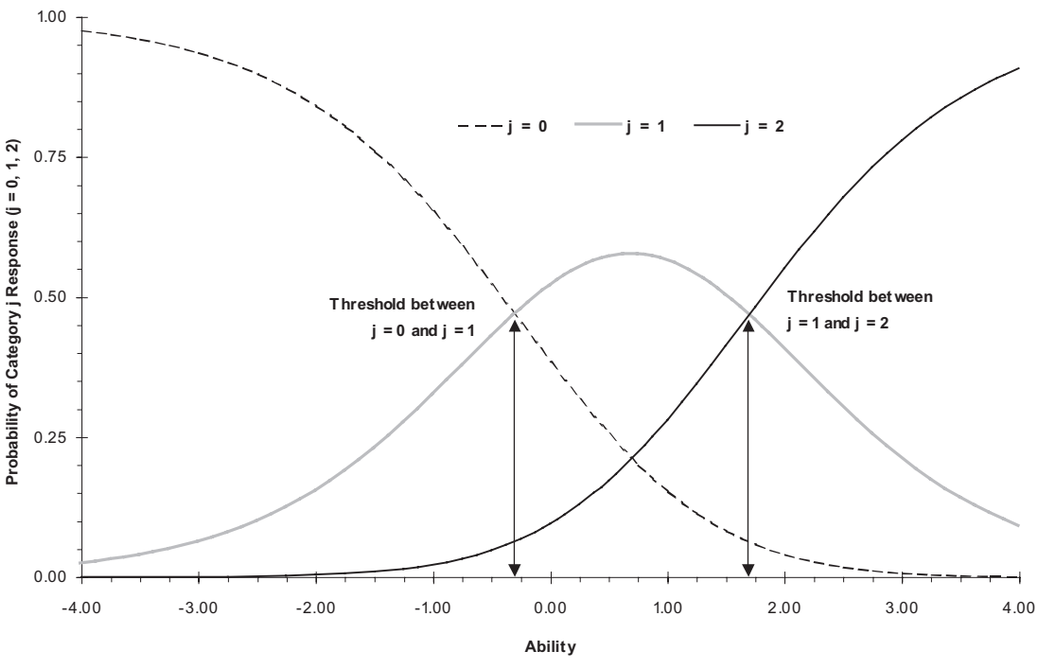

Figure 6.3: Category Response Curves for a Two-step Item

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 (marked by the right arrow), 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 $x = 0, 1, ..., m_i$, and by definition,

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

The above equation gives the probability of scoring x on the i-th test item as a function of ability (θ) and the difficulty 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

A Forms Field Test of the AZELLA was conducted in the spring of 2006. The testing window was between February 20 and March 10, 2006. Calibration, equating, and scaling were conducted using data collected from this testing period. Harcourt Assessment 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, were fixed to the parameter values from the pre-existing vertical scale for the 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 and the SELP item bank. Any remaining non-SELP items on the AZELLA 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. 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 and
- 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	23	14
Primary 1–2	76	100	43	35
Elementary 3–5	76	100	36	31
Middle 6–8	84	110	42	33
High School 9–12	84	110	35	28

The scale was then obtained by taking the item parameters for Total Composite and using it to create raw score-to-scaled score tables. Finally, when these calibrations and scales were completed, the Forms Field Test items for the 2006 administration 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.

6.3 Vertical Scaling of Preliteracy Level to the Primary Level

The Preliteracy level of the AZELLA was not on the SELP vertical scale. A separate study had to be conducted in order to place the Preliteracy level of the AZELLA on the new AZELLA scale. An important component of any multilevel test is a continuous score scale that permits the interpretation of scores across levels of the test. According

to Nitko (2004), a vertical scale is defined as an extended score scale that spans a series of levels and allows for the estimation of student growth along a continuum. In conducting the AZELLA Preliteracy/Primary multilevel equating, the adjacent levels of the test were scaled so that scores across levels could be expressed on the same scale. The design that was utilized to obtain the vertical scale for the Preliteracy level of the AZELLA was the common-person linking design, which is also referred to as the equivalent groups design (Kolen & Brennan, 2004, p. 389).

To accomplish the vertical scaling process, students in grades K and 1 were recruited for this study. In the common-person linking design, the same students were administered both the Preliteracy and Primary levels of the test. In addition, some students took only the Preliteracy level of the test or only the Primary level. The dataset was combined and concurrent calibrations were conducted on the students from grades K and 1. In order to be placed on the same AZELLA vertical scale, the anchor items were fixed to the parameter values from the Primary level calibration conducted previously. The remaining Preliteracy items were calibrated together with the Primary items using the Rasch and Partial Credit models. Fixing the values of the Primary anchor items prior to calibration resulted in the item difficulty and step parameters of all the items being placed on the same ability metric. The results of this calibration were then used as the operational item parameters to create the scales for the Preliteracy AZELLA.

6.4 Scaled Scores

The following equation was used to derive the AZELLA scaled scores:

$$SS = 35*(theta) + 600$$

The AZELLA scaling procedure involves linear transformations of the raw score points into scaled score points. 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), Harcourt Assessment also analyzed the items on the AZELLA 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 assessment. These models were used for several reasons.

First, the AZELLA vertical scale was created based on the pre-existing SELP vertical scale that was developed using the Rasch model. By using SELP items with known Rasch item difficulties, Harcourt Assessment 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 Harcourt Assessment 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 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). None of the items were flagged for Infit. The percentage of items that were flagged for Outfit varied depending on grade spans. In general, around 1%–11% 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 Information

Table 7.1 presents the number of items in each subtest per grade span, the maximum number of points attainable for each subtest, and the average Rasch difficulty for each subtest.

Table 7.1: Average Rasch Difficulty by Grade Span by Subtest

Level	Test	# of Items	Max Points	Average Rasch Difficulty
Preliteracy	Listening	12	12	–3.07
K	Prereading	12	12	–2.31
	Prewriting	17	22	–2.06
	Speaking	12	26	–2.27
	Total	53	72	–2.39
Primary	Listening	20	20	–2.42
1–2	Writing Conventions	20	20	0.11
	Reading	20	20	–0.37
	Writing	2	8	0.29
	Speaking	14	32	–1.06
	Total	76	100	–0.89
Elementary	Listening	20	20	0.06
3–5	Writing Conventions	20	20	0.02
	Reading	20	20	0.74
	Writing	2	8	0.96
	Speaking	14	32	–0.05
	Total	76	100	0.23
Middle	Listening	20	20	0.51
6–8	Writing Conventions	24	24	0.55
	Reading	24	24	1.09
	Writing	2	8	0.79
	Speaking	14	34	0.35
	Total	84	110	0.67
High School	Listening	20	20	1.40
9–12	Writing Conventions	24	24	1.24
	Reading	24	24	1.64
	Writing	2	8	1.96
	Speaking	14	34	0.96
	Total	84	110	1.36

Appendices A.1–A.5 contain the results of the operational 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),
- 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, and
- 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

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 of the AZELLA. Sections 8.2 through 8.9 provide descriptions of the AZELLA 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 is 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, and
- balanced regional representation.

The AZELLA panelists were recruited by the ADE to participate in the standard setting meeting. They were placed in one of the five committees according to their experience in teaching at the various levels.

Five separate standard setting committees 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

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

8.4 The Standard Setting Process

Harcourt Assessment recommended a modified-Angoff procedure (Angoff, 1984) for producing the suggested cut scores for the AZELLA. 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 fifteen 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 is 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 in the technical documentation. The judges were told to feel free to raise questions during the sessions. Although an important goal of the process is for judges to approach consensus or convergence in ratings, it is 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. Fitzpatrick suggests that conformity due to social pressure is not desirable in standard setting. Items with disparate ratings will be discussed in order to educate the judges about other judges' rationales behind their ratings. Any potential effects of undue social pressure will be moderated through the group process skills of the leader of 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 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.

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 levels 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 8.2a and 8.2b provide the approved AZELLA cut score ranges in scaled score for the reporting strands for all grades.

Table 8.2a: 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
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 8.2b: 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
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. Administration Results

This section presents the results of the Wave 1 AZELLA 2006 administration (Operational). Analyses are provided for all the reporting strands. The following are the reporting strands:

- Listening,
- Speaking,
- Comprehension (Listening + Reading),
- Oral (Listening + Speaking),
- Reading,
- 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 items,
- number of students,
- means,
- median,
- inter quartile range (IQR), and
- standard deviations.

Table 9.1: Percentage of Students in Each Proficiency Category

			Proficiency Level Percentages				
Grade	Strand	N-Count	Pre-Emergent	Emergent	Basic	Intermediate	Proficient
K	Listening	28482	4.32	3.61	14.17	33.73	44.17
	Prereading	28482	15.20	16.69	39.57	19.53	9.01
	Speaking	28482	6.71	6.28	32.93	26.32	27.76
	Comprehension	28482	5.67	5.60	45.99	33.49	9.25
	Oral	28482	4.29	6.11	32.85	32.55	24.19
	Prewriting	28482	29.29	11.92	49.30	8.86	0.63
	Total	28482	7.18	9.99	53.61	26.84	2.39
1	Listening	8987	2.60	0.42	5.31	33.79	57.87
	Reading	8987	13.39	11.17	59.00	13.98	2.47
	Speaking	8987	13.07	2.86	19.45	31.95	32.67
	Comprehension	8987	2.95	1.13	42.23	49.49	4.19
	Oral	8987	4.51	4.01	18.54	39.06	33.89
	Total Writing	8987	21.04	14.21	57.41	6.50	0.85
	Total	8987	5.51	5.39	56.21	31.02	1.87
2	Listening	7013	2.40	0.37	4.19	24.51	68.53
	Reading	7013	5.02	4.11	39.51	28.89	22.47
	Speaking	7013	9.74	1.25	9.62	29.63	49.75
	Comprehension	7013	2.55	0.97	14.94	54.46	27.08
	Oral	7013	4.33	2.67	7.93	33.14	51.93
	Total Writing	7013	5.99	6.59	37.90	40.74	8.78
	Total	7013	4.09	3.21	19.12	57.32	16.26
3	Listening	5302	5.30	2.87	28.97	43.10	19.77
	Reading	5302	7.64	7.36	46.93	27.80	10.28
	Speaking	5302	12.05	1.64	7.36	23.31	55.64
	Comprehension	5302	5.66	3.60	43.23	37.40	10.11
	Oral	5302	8.07	3.96	10.90	39.34	37.72
	Total Writing	5302	8.36	7.54	29.12	34.78	20.20
	Total	5302	6.83	4.87	23.26	49.77	15.28
4	Listening	4939	4.54	1.96	16.93	38.57	38.00
	Reading	4939	5.99	4.52	35.25	32.54	21.70
	Speaking	4939	11.97	1.50	6.13	22.86	57.54
	Comprehension	4939	5.06	1.92	31.32	38.10	23.59
	Oral	4939	9.07	2.55	9.68	30.73	47.97
	Total Writing	4939	6.60	5.65	18.02	32.25	37.48
	Total	4939	7.07	4.09	16.72	45.98	26.14
5	Listening	4592	5.64	1.98	15.44	26.66	50.28
	Reading	4592	7.27	5.12	28.85	26.46	32.30
	Speaking	4592	12.37	1.11	6.16	18.58	61.78
	Comprehension	4592	5.27	3.70	23.91	36.54	30.57
	Oral	4592	10.13	1.87	7.56	28.07	52.37
	Total Writing	4592	6.21	4.22	13.83	35.00	40.74
	Total	4592	8.01	3.48	13.20	41.31	33.99

Table 9.1: Percentage of Students in Each Proficiency Category *(continued)*

			Proficiency Level Percentages				
Grade	**Strand**	**N-Count**	**Pre-Emergent**	**Emergent**	**Basic**	**Intermediate**	**Proficient**
6	Listening	4089	4.40	2.49	19.86	45.24	28.00
	Reading	4089	5.14	4.65	32.55	33.70	23.97
	Speaking	4089	10.56	1.44	4.70	16.48	66.81
	Comprehension	4089	4.30	2.96	29.71	41.60	21.42
	Oral	4089	7.21	3.15	7.29	29.54	52.80
	Total Writing	4089	4.92	2.89	18.32	40.18	33.70
	Total	4089	5.01	4.26	14.60	43.73	32.40
7	Listening	4190	5.32	2.32	16.66	38.26	37.45
	Reading	4190	6.59	4.06	26.47	32.29	30.60
	Speaking	4190	12.91	2.22	6.32	18.69	59.86
	Comprehension	4190	5.66	3.84	24.94	40.21	25.35
	Oral	4190	9.16	3.79	9.45	27.21	50.38
	Total Writing	4190	5.42	2.43	18.97	32.55	40.62
	Total	4190	6.56	4.80	15.32	40.74	32.58
8	Listening	3815	4.30	2.80	16.67	43.75	32.48
	Reading	3815	5.03	5.71	22.80	35.47	30.98
	Speaking	3815	11.35	2.15	7.31	21.78	57.40
	Comprehension	3815	4.40	3.70	21.70	43.56	26.63
	Oral	3815	7.21	4.06	9.93	32.84	45.95
	Total Writing	3815	3.91	2.39	17.35	37.46	38.90
	Total	3815	5.50	4.01	14.60	45.56	30.33
9	Listening	5179	2.90	2.49	26.36	45.20	23.05
	Reading	5179	3.44	4.33	29.21	36.69	26.34
	Speaking	5179	11.99	1.97	6.18	15.91	63.95
	Comprehension	5179	2.80	3.15	29.16	44.56	20.33
	Oral	5179	6.78	5.27	9.23	29.87	48.85
	Total Writing	5179	2.70	4.25	18.05	37.52	37.48
	Total	5179	3.69	5.91	15.45	43.19	31.76
10	Listening	2849	2.81	2.67	28.08	43.42	23.03
	Reading	2849	2.67	4.53	28.22	34.47	30.12
	Speaking	2849	12.85	2.28	9.41	14.36	61.11
	Comprehension	2849	2.56	3.09	32.61	43.31	18.43
	Oral	2849	6.60	6.53	12.21	32.08	42.58
	Total Writing	2849	2.88	3.90	22.43	41.80	28.99
	Total	2849	4.04	5.86	18.57	43.88	27.66
11	Listening	1962	1.17	1.58	22.78	44.55	29.92
	Reading	1962	1.38	1.78	22.27	43.07	31.50
	Speaking	1962	7.90	1.63	8.56	18.20	63.71
	Comprehension	1962	1.12	1.89	26.40	43.93	26.66
	Oral	1962	3.77	4.18	11.72	33.84	46.48
	Total Writing	1962	0.92	2.80	17.84	42.71	35.73
	Total	1962	1.89	3.36	17.18	46.59	30.99
12	Listening	1768	1.75	3.17	26.36	50.51	18.21
	Reading	1768	1.47	4.64	23.08	38.01	32.81
	Speaking	1768	7.13	2.38	9.79	22.12	58.60
	Comprehension	1768	1.02	3.96	24.77	50.23	20.02
	Oral	1768	3.68	4.24	13.69	36.82	41.57
	Total Writing	1768	2.04	2.77	20.14	45.81	29.24
	Total	1768	2.04	3.62	18.78	49.38	26.19

Table 9.2: Raw Score Descriptive Statistics by Grade

Grade	Strand	N-Count	Mean	Median	St. Dev.	IQR
K	Listening	28482	7.54	8	3.02	4
	Prereading	28482	4.05	4	2.93	4
	Speaking	28482	13.99	15	7.23	12
	Comprehension	28482	11.59	12	5.11	6
	Oral	28482	21.53	23	9.49	15
	Prewriting	28482	5.82	6	4.28	7
	Total	28482	31.40	32	13.78	19
1	Listening	8987	15.06	16	3.59	3
	Reading	8987	7.09	7	3.88	4
	Speaking	8987	19.29	22	9.71	13
	Comprehension	8987	22.15	23	6.29	7
	Oral	8987	34.35	38	12.13	16
	Total Writing	8987	8.25	8	4.76	6
	Total	8987	49.68	52	17.38	20
2	Listening	7013	16.55	17	3.52	2
	Reading	7013	11.90	12	4.93	7
	Speaking	7013	23.25	26	9.24	10
	Comprehension	7013	28.45	29	7.48	9
	Oral	7013	39.81	44	11.80	11
	Total Writing	7013	14.79	15	6.06	8
	Total	7013	66.50	71	20.14	22
3	Listening	5302	11.49	12	4.55	6
	Reading	5302	9.17	9	4.63	7
	Speaking	5302	22.15	26	9.92	11
	Comprehension	5302	20.66	21	8.49	12
	Oral	5302	33.64	38	13.34	14
	Total Writing	5302	15.27	17	6.85	11
	Total	5302	58.08	63	22.66	27
4	Listening	4939	13.26	14	4.71	6
	Reading	4939	10.89	11	5.01	8
	Speaking	4939	23.09	27	10.00	10
	Comprehension	4939	24.15	25	9.05	12
	Oral	4939	36.34	41	13.66	14
	Total Writing	4939	17.67	20	7.04	10
	Total	4939	64.90	72	23.64	26
5	Listening	4592	14.08	16	4.91	6
	Reading	4592	12.09	13	5.28	8
	Speaking	4592	23.70	28	10.08	9
	Comprehension	4592	26.17	28	9.60	12
	Oral	4592	37.78	43	14.11	12
	Total Writing	4592	19.06	21	7.20	8
	Total	4592	68.93	77	24.85	27

Table 9.2: Raw Score Descriptive Statistics by Grade *(continued)*

Grade	Strand	N-Count	Mean	Median	St. Dev.	IQR
6	Listening	4089	12.53	13	4.53	6
	Reading	4089	13.45	14	6.00	9
	Speaking	4089	25.54	30	10.39	10
	Comprehension	4089	25.99	27	9.89	14
	Oral	4089	38.07	43	13.93	12
	Total Writing	4089	20.72	23	7.82	11
	Total	4089	72.24	79	25.59	29
7	Listening	4190	13.05	14	4.88	6
	Reading	4190	14.16	15	6.34	10
	Speaking	4190	24.61	29	11.13	12
	Comprehension	4190	27.20	29	10.64	15
	Oral	4190	37.66	44	15.05	15
	Total Writing	4190	21.38	24	8.17	11
	Total	4190	73.19	82	27.66	33
8	Listening	3815	13.81	15	4.60	5
	Reading	3815	15.33	17	6.15	9
	Speaking	3815	25.34	30	10.65	11
	Comprehension	3815	29.14	32	10.17	14
	Oral	3815	39.14	45	14.13	14
	Total Writing	3815	22.61	25	7.49	9
	Total	3815	77.09	86	25.70	30
9	Listening	5179	12.13	13	4.20	5
	Reading	5179	14.13	15	5.71	9
	Speaking	5179	24.81	29	10.92	11
	Comprehension	5179	26.26	28	9.30	14
	Oral	5179	36.93	42	14.10	15
	Total Writing	5179	20.67	23	7.32	10
	Total	5179	71.73	80	25.30	30
10	Listening	2849	12.02	13	4.29	6
	Reading	2849	14.50	15	5.75	9
	Speaking	2849	24.17	29	11.38	14
	Comprehension	2849	26.53	28	9.40	14
	Oral	2849	36.20	42	14.47	18
	Total Writing	2849	20.51	22	7.26	10
	Total	2849	71.21	79	25.62	32
11	Listening	1962	12.97	14	3.92	6
	Reading	1962	16.02	17	5.31	8
	Speaking	1962	26.29	30	9.83	11
	Comprehension	1962	28.99	31	8.59	13
	Oral	1962	39.26	44	12.49	14
	Total Writing	1962	22.11	24	6.33	8
	Total	1962	77.39	84	22.19	28
12	Listening	1768	12.90	13	3.95	6
	Reading	1768	16.12	17	5.32	7
	Speaking	1768	26.30	30	9.59	10
	Comprehension	1768	29.02	31	8.60	12
	Oral	1768	39.20	44	12.28	14
	Total Writing	1768	22.08	24	6.53	8
	Total	1768	77.40	84	22.07	28

Table 9.3: Scaled Score Descriptive Statistics by Grade

Grade	Strand	N-Count	Mean	Median	St. Dev.	IQR
K	Listening	28482	518.60	522	61.49	65
	Prereading	28482	481.26	491	59.97	62
	Speaking	28482	530.89	539	69.37	83
	Comprehension	28482	500.23	506	52.31	42
	Oral	28482	527.50	534	60.01	73
	Prewriting	28482	465.23	482	73.69	100
	Total	28482	506.42	512	48.58	53
1	Listening	8987	577.43	586	55.75	48
	Reading	8987	555.05	563	51.39	37
	Speaking	8987	578.14	589	75.12	60
	Comprehension	8987	564.02	571	46.58	39
	Oral	8987	578.03	588	57.20	57
	Total Writing	8987	561.14	569	46.84	43
	Total	8987	568.47	576	45.99	34
2	Listening	7013	607.19	604	61.00	40
	Reading	7013	607.95	605	60.94	62
	Speaking	7013	607.88	611	78.69	72
	Comprehension	7013	604.54	604	57.41	55
	Oral	7013	604.68	614	61.92	58
	Total Writing	7013	608.22	611	48.31	48
	Total	7013	603.34	611	53.64	46
3	Listening	5302	615.01	620	54.10	52
	Reading	5302	615.59	620	57.34	63
	Speaking	5302	637.59	650	89.41	86
	Comprehension	5302	615.03	620	53.11	53
	Oral	5302	624.57	637	61.44	50
	Total Writing	5302	614.89	626	60.73	73
	Total	5302	618.49	628	56.98	48
4	Listening	4939	634.97	637	58.31	60
	Reading	4939	634.38	637	61.64	72
	Speaking	4939	647.55	658	92.85	82
	Comprehension	4939	633.76	637	56.26	56
	Oral	4939	638.64	649	64.83	59
	Total Writing	4939	635.84	648	64.88	75
	Total	4939	634.24	646	58.80	53
5	Listening	4592	645.43	658	64.16	69
	Reading	4592	648.25	654	68.04	74
	Speaking	4592	654.54	668	95.80	100
	Comprehension	4592	645.39	651	63.28	59
	Oral	4592	647.05	658	70.68	58
	Total Writing	4592	650.44	656	71.96	69
	Total	4592	644.66	657	65.95	62

Table 9.3: Scaled Score Descriptive Statistics by Grade *(continued)*

Grade	Strand	N-Count	Mean	Median	St. Dev.	IQR
6	Listening	4089	643.82	647	57.46	58
	Reading	4089	649.75	654	61.28	65
	Speaking	4089	670.21	681	87.24	98
	Comprehension	4089	646.21	651	57.11	59
	Oral	4089	653.12	665	62.36	49
	Total Writing	4089	649.87	660	62.68	68
	Total	4089	649.41	659	58.39	52
7	Listening	4190	650.11	656	63.89	63
	Reading	4190	656.31	661	67.05	75
	Speaking	4190	664.27	671	92.78	107
	Comprehension	4190	652.21	659	62.88	66
	Oral	4190	653.00	669	69.00	61
	Total Writing	4190	655.55	666	67.60	73
	Total	4190	651.85	664	64.81	62
8	Listening	3815	660.39	667	60.46	54
	Reading	3815	668.12	676	64.60	71
	Speaking	3815	671.23	681	89.90	103
	Comprehension	3815	663.27	672	59.40	64
	Oral	3815	661.52	674	63.70	64
	Total Writing	3815	665.36	674	61.84	63
	Total	3815	661.43	672	57.54	61
9	Listening	5179	670.15	677	48.03	45
	Reading	5179	675.34	680	55.42	67
	Speaking	5179	687.87	695	91.72	99
	Comprehension	5179	672.16	679	48.35	57
	Oral	5179	674.94	687	57.36	57
	Total Writing	5179	680.47	692	56.64	64
	Total	5179	675.03	686	52.39	53
10	Listening	2849	669.41	677	49.09	54
	Reading	2849	679.19	680	55.50	67
	Speaking	2849	685.37	695	95.53	112
	Comprehension	2849	674.13	679	47.75	57
	Oral	2849	673.10	687	58.42	65
	Total Writing	2849	679.56	685	56.10	64
	Total	2849	674.84	685	52.03	56
11	Listening	1962	680.15	686	43.92	57
	Reading	1962	694.08	695	52.15	63
	Speaking	1962	705.03	703	85.20	144
	Comprehension	1962	686.81	691	43.08	57
	Oral	1962	686.51	695	50.77	58
	Total Writing	1962	692.28	699	48.32	56
	Total	1962	687.98	694	44.44	55
12	Listening	1768	679.27	677	44.65	57
	Reading	1768	694.57	695	51.09	56
	Speaking	1768	704.29	703	82.39	95
	Comprehension	1768	686.36	691	43.16	50
	Oral	1768	686.14	695	49.91	58
	Total Writing	1768	691.99	699	49.40	56
	Total	1768	687.74	694	43.78	55

Arizona English Language Learner Assessment (AZELLA)

Appendices A–E

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	In Fit	Out Fit
1	MC	1	317	0.97	0.30	–4.74	0.29	0.78	0.56
2	MC	1	317	0.89	0.44	–3.56	0.19	0.89	0.69
3	MC	1	317	0.95	0.32	–4.66	0.28	1.00	0.72
4	MC	1	317	0.83	0.50	–2.98	0.16	0.87	0.68
5	MC	1	317	0.91	0.29	–3.93	0.22	1.07	0.88
6	MC	1	317	0.84	0.55	–3.13	0.17	0.84	0.64
7	MC	1	317	0.77	0.56	–2.67	0.15	0.89	0.81
8	MC	1	317	0.71	0.28	–2.25	0.14	1.17	1.20
9	MC	1	317	0.76	0.25	–2.55	0.15	1.18	1.72
10	MC	1	317	0.75	0.34	–2.46	0.14	1.09	1.14
11	MC	1	317	0.77	0.32	–2.55	0.15	1.09	1.23
12	MC	1	317	0.58	0.25	–1.40	0.13	1.18	1.27

Prereading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
13	MC	1	317	0.84	0.34	–2.85	0.16	0.95	0.88
14	MC	1	317	0.70	0.56	–2.18	0.14	0.87	0.83
15	MC	1	317	0.40	0.18	–0.45	0.13	1.22	1.46
16	MC	1	317	0.50	0.26	–0.92	0.12	1.14	1.34
17	MC	1	317	0.77	0.47	–2.61	0.15	0.98	0.84
18	MC	1	317	0.64	0.40	–1.82	0.13	1.04	0.99
19	MC	1	317	0.73	0.44	–2.23	0.14	0.97	0.91
20	MC	1	317	0.83	0.41	–3.01	0.16	1.03	0.82
21	MC	1	317	0.84	0.50	–3.08	0.17	0.91	0.73
22	MC	1	317	0.82	0.45	–2.96	0.16	0.95	0.83
23	MC	1	317	0.81	0.41	–2.83	0.16	0.99	0.88
24	MC	1	317	0.80	0.42	–2.75	0.15	1.00	0.87

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
25	SS	2	317	1.92	0.35	–4.00	0.19	1.12	0.70
26	SS	2	317	1.90	0.43	–3.94	0.18	1.01	0.67
27	SS	2	317	1.81	0.35	–3.61	0.14	1.02	0.92
28	SS	2	317	1.69	0.52	–3.33	0.12	0.90	0.88
29	SS	2	317	1.29	0.58	–1.74	0.09	0.93	0.91
30	SS	2	317	1.33	0.65	–1.78	0.09	0.89	0.87
31	SE	2	317	1.30	0.67	–1.73	0.09	0.81	0.78
32	SE	4	317	2.12	0.69	–1.09	0.06	0.98	0.94
33	SE	2	317	1.34	0.63	–1.77	0.09	0.85	0.94
34	SS	2	317	1.21	0.61	–1.51	0.08	0.97	0.90
35	SS	2	317	1.01	0.60	–1.05	0.08	0.93	0.87
36	SS	2	317	1.28	0.57	–1.68	0.09	1.01	1.00

Prewriting

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
37	WR	1	317	0.97	0.17	–4.66	0.28	0.74	1.00
38	WR	1	317	0.93	0.16	–4.17	0.23	1.12	1.56
39	WR	1	317	0.93	0.17	–4.12	0.23	1.06	1.58
40	WR	1	317	0.91	0.16	–3.70	0.20	1.13	1.20
41	WR	2	317	1.10	0.30	–1.31	0.09	1.40	1.42
42	WR	2	317	1.19	0.34	–1.48	0.09	1.40	1.40
43	WR	1	317	0.88	0.49	–3.56	0.19	0.88	0.73
44	WR	1	317	0.58	0.38	–1.54	0.13	1.05	1.03
45	WR	1	317	0.95	0.32	–4.59	0.27	0.96	0.93
46	WR	1	317	0.79	0.50	–2.63	0.15	0.87	0.83
47	WR	1	317	0.75	0.52	–2.41	0.14	0.88	0.84
48	WR	1	317	0.80	0.48	–2.77	0.15	0.91	0.97
49	WR	1	317	0.47	0.50	–1.14	0.12	0.88	0.87
50	WR	1	317	0.37	0.46	–0.67	0.13	0.91	0.83
51	WR	2	317	0.21	0.39	1.27	0.13	0.72	0.71
52	WR	2	317	0.11	0.32	1.84	0.15	0.68	0.55
53	WR	2	317	0.29	0.45	0.64	0.11	0.70	0.61

A.2: Primary

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
1	MC	1	554	0.99	0.27	–4.60	0.40	1.01	0.25
2	MC	1	554	0.97	0.34	–3.48	0.24	0.89	0.56
3	MC	1	554	0.97	0.36	–3.60	0.25	0.84	0.70
4	MC	1	554	0.96	0.21	–3.32	0.22	1.00	1.14
5	MC	1	554	0.97	0.37	–3.30	0.22	0.71	0.30
6	MC	1	554	0.98	0.39	–3.97	0.30	0.82	0.24
7	MC	1	554	0.98	0.38	–3.89	0.29	0.83	0.28
8	MC	1	554	0.98	0.13	–3.73	0.27	0.87	1.18
9	MC	1	554	0.55	0.16	0.04	0.09	1.24	1.37
10	MC	1	554	0.95	0.22	–3.18	0.21	1.03	1.02
11	MC	1	554	0.97	0.20	–3.09	0.20	0.77	0.54
12	MC	1	554	0.92	0.34	–2.19	0.15	0.71	0.79
13	MC	1	554	0.93	0.25	–2.73	0.18	1.09	0.98
14	MC	1	554	0.92	0.42	–2.58	0.17	0.89	0.58
15	MC	1	554	0.92	0.25	–2.91	0.19	1.32	1.47
16	MC	1	554	0.70	0.20	–0.64	0.10	1.15	1.22
17	MC	1	554	0.80	0.22	–1.69	0.13	1.35	1.66
18	MC	1	554	0.59	0.33	–0.15	0.09	1.06	1.07
19	MC	1	554	0.60	0.33	–0.19	0.09	1.07	1.10
20	MC	1	554	0.38	0.25	0.87	0.10	1.14	1.20

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
21	MC	1	554	0.23	0.16	1.73	0.11	1.12	1.56
22	MC	1	554	0.84	0.45	–1.34	0.11	0.76	0.65
23	MC	1	554	0.66	0.54	–0.37	0.10	0.83	0.77
24	MC	1	554	0.66	0.47	–0.66	0.10	0.98	0.90
25	MC	1	554	0.66	0.48	–0.47	0.10	0.92	0.83
26	MC	1	554	0.60	0.52	–0.02	0.09	0.86	0.83
27	MC	1	554	0.55	0.48	–0.43	0.10	1.05	1.06
28	MC	1	554	0.80	0.47	–1.24	0.11	0.89	0.72
29	MC	1	554	0.42	0.23	0.43	0.09	1.15	1.17
30	MC	1	554	0.67	0.44	–0.58	0.10	0.96	0.88
31	MC	1	554	0.57	0.39	–0.05	0.09	0.99	1.00
32	MC	1	554	0.56	0.35	–0.01	0.09	1.04	1.04
33	MC	1	554	0.58	0.44	–0.11	0.09	0.96	0.91
34	MC	1	554	0.62	0.47	–0.31	0.10	0.93	0.86
35	MC	1	554	0.38	0.14	0.87	0.10	1.19	1.43
36	MC	1	554	0.49	0.18	0.37	0.09	1.18	1.38
37	MC	1	554	0.37	0.11	0.92	0.10	1.27	1.49
38	MC	1	554	0.29	0.05	1.36	0.10	1.26	1.72
39	MC	1	554	0.44	0.16	0.62	0.09	1.22	1.43
40	MC	1	554	0.28	0.13	1.44	0.10	1.20	1.66

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
41	MC	1	554	0.94	0.29	–2.52	0.16	0.72	0.58
42	MC	1	554	0.83	0.42	–1.42	0.12	0.86	0.81
43	MC	1	554	0.74	0.38	–1.24	0.11	1.18	1.01
44	MC	1	554	0.89	0.38	–2.01	0.14	0.87	0.65
45	MC	1	554	0.63	0.48	–0.31	0.10	0.91	0.83
46	MC	1	554	0.75	0.38	–1.18	0.11	1.08	0.96
47	MC	1	554	0.68	0.38	–0.58	0.10	1.01	1.02
48	MC	1	554	0.57	0.55	–0.05	0.09	0.84	0.80
49	MC	1	554	0.58	0.32	–0.38	0.10	1.15	1.17
50	MC	1	554	0.51	0.34	0.24	0.09	1.04	1.07
51	MC	1	554	0.52	0.33	–0.18	0.09	1.13	1.12
52	MC	1	554	0.54	0.53	–0.19	0.09	0.91	0.86
53	MC	1	554	0.53	0.25	–0.17	0.09	1.19	1.21
54	MC	1	554	0.35	0.37	0.67	0.09	0.94	1.00
55	MC	1	554	0.53	0.41	0.16	0.09	0.96	0.96
56	MC	1	554	0.48	0.43	0.41	0.09	0.93	1.01
57	MC	1	554	0.57	0.47	–0.03	0.09	0.92	0.88
58	MC	1	554	0.47	0.42	0.46	0.09	0.95	0.99
59	MC	1	554	0.49	0.24	0.36	0.09	1.12	1.25
60	MC	1	554	0.47	0.33	0.47	0.09	1.02	1.15

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
61	ER	4	554	1.99	0.67	0.19	0.05	0.75	0.76
62	ER	4	554	2.04	0.73	0.38	0.05	0.71	0.72

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
63	SS	2	554	1.66	0.45	–1.14	0.08	1.07	1.32
64	SS	2	554	1.62	0.58	–1.11	0.08	0.86	0.97
65	SS	2	554	1.60	0.64	–1.18	0.08	0.80	0.75
66	SS	2	554	1.62	0.65	–1.14	0.08	0.80	0.78
67	SS	2	554	1.68	0.58	–1.27	0.08	0.72	0.68
68	SS	2	554	1.63	0.57	–1.26	0.08	0.80	0.75
69	SS	2	554	1.63	0.62	–1.22	0.08	0.75	0.70
70	SS	2	554	1.59	0.60	–1.07	0.08	0.75	0.72
71	SS	2	554	1.69	0.59	–1.46	0.08	0.73	0.76
72	SS	4	554	3.04	0.68	–0.72	0.05	1.01	0.99
73	SS	4	554	2.74	0.64	–0.37	0.05	1.08	1.21
74	SS	2	554	1.48	0.56	–0.72	0.07	0.93	1.05
75	SS	2	554	1.63	0.61	–1.12	0.08	0.81	0.73
76	SS	2	554	1.58	0.57	–1.00	0.07	0.88	0.90

A.3: Elementary

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
1	MC	1	463	0.89	0.45	–1.45	0.17	1.04	1.09
2	MC	1	463	0.74	0.58	0.24	0.11	0.78	0.75
3	MC	1	463	0.94	0.45	–2.03	0.20	0.89	0.42
4	MC	1	463	0.67	0.43	0.20	0.11	1.10	1.17
5	MC	1	463	0.82	0.54	–0.60	0.13	0.87	0.83
6	MC	1	463	0.78	0.44	–0.19	0.12	0.98	0.95
7	MC	1	463	0.68	0.46	0.33	0.11	1.00	0.97
8	MC	1	463	0.68	0.42	0.05	0.12	1.16	1.18
9	MC	1	463	0.57	0.35	0.95	0.11	1.11	1.19
10	MC	1	463	0.78	0.67	–0.32	0.13	0.74	0.58
11	MC	1	463	0.49	0.29	1.32	0.10	1.16	1.35
12	MC	1	463	0.52	0.32	1.01	0.11	1.16	1.26
13	MC	1	463	0.90	0.15	–1.40	0.17	1.20	1.88
14	MC	1	463	0.62	0.45	0.68	0.11	1.01	1.01
15	MC	1	463	0.76	0.26	–0.20	0.12	1.20	1.53
16	MC	1	463	0.68	0.38	0.32	0.11	1.10	1.13
17	MC	1	463	0.79	0.32	–0.42	0.13	1.13	1.61
18	MC	1	463	0.52	0.29	1.18	0.10	1.17	1.32
19	MC	1	463	0.52	0.35	1.18	0.10	1.10	1.17
20	MC	1	463	0.68	0.22	0.35	0.11	1.29	1.41

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
21	MC	1	463	0.92	0.29	–1.70	0.18	1.06	1.14
22	MC	1	463	0.92	0.33	–1.80	0.19	1.01	0.87
23	MC	1	463	0.93	0.43	–1.84	0.19	0.92	0.54
24	MC	1	463	0.82	0.49	–0.66	0.14	0.96	0.76
25	MC	1	463	0.70	0.46	0.48	0.11	0.93	0.90
26	MC	1	463	0.63	0.49	0.70	0.11	0.94	0.91
27	MC	1	463	0.72	0.50	–0.14	0.12	1.05	1.06
28	MC	1	463	0.66	0.51	0.45	0.11	0.95	0.96
29	MC	1	463	0.67	0.51	0.13	0.12	1.02	0.98
30	MC	1	463	0.80	0.53	–0.46	0.13	0.91	0.76
31	MC	1	463	0.63	0.48	0.43	0.11	1.01	1.08
32	MC	1	463	0.61	0.35	0.82	0.11	1.09	1.18
33	MC	1	463	0.51	0.32	1.01	0.11	1.15	1.26
34	MC	1	463	0.81	0.55	–0.55	0.13	0.90	0.67
35	MC	1	463	0.78	0.51	–0.32	0.13	0.93	0.82
36	MC	1	463	0.65	0.55	0.49	0.11	0.88	0.83
37	MC	1	463	0.55	0.38	1.03	0.11	1.07	1.09
38	MC	1	463	0.48	0.32	1.40	0.10	1.11	1.27
39	MC	1	463	0.76	0.51	–0.18	0.12	0.94	0.84
40	MC	1	463	0.53	0.36	1.12	0.10	1.09	1.20

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
41	MC	1	463	0.89	0.31	–1.74	0.18	1.38	1.58
42	MC	1	463	0.90	0.45	–1.43	0.17	0.90	0.85
43	MC	1	463	0.86	0.49	–1.02	0.15	0.90	0.88
44	MC	1	463	0.70	0.43	0.23	0.11	1.03	1.05
45	MC	1	463	0.55	0.40	1.01	0.11	1.06	1.07
46	MC	1	463	0.56	0.49	0.97	0.11	0.93	0.95
47	MC	1	463	0.48	0.26	1.39	0.10	1.21	1.32
48	MC	1	463	0.70	0.49	0.07	0.12	1.02	0.93
49	MC	1	463	0.65	0.57	0.28	0.11	0.94	0.86
50	MC	1	463	0.56	0.51	0.97	0.11	0.92	0.90
51	MC	1	463	0.51	0.45	1.24	0.10	0.99	1.00
52	MC	1	463	0.53	0.54	0.82	0.11	0.93	0.89
53	MC	1	463	0.60	0.55	0.49	0.11	0.95	0.90
54	MC	1	463	0.28	0.16	2.42	0.11	1.14	1.76
55	MC	1	463	0.49	0.34	1.18	0.10	1.12	1.17
56	MC	1	463	0.51	0.49	0.98	0.11	0.95	0.97
57	MC	1	463	0.47	0.43	1.44	0.10	0.96	1.06
58	MC	1	463	0.34	0.10	2.10	0.11	1.28	1.77
59	MC	1	463	0.46	0.37	1.50	0.10	1.07	1.16
60	MC	1	463	0.38	0.23	1.91	0.11	1.19	1.36

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
61	ER	4	463	2.51	0.76	1.02	0.06	1.08	1.12
62	ER	4	463	2.49	0.74	0.90	0.06	0.97	1.04

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
63	SS	2	463	1.61	0.60	–0.25	0.09	0.98	1.07
64	SS	2	463	1.33	0.67	0.42	0.08	0.87	0.83
65	SS	2	463	1.48	0.61	–0.09	0.09	0.93	0.93
66	SS	2	463	1.44	0.62	–0.01	0.09	0.93	0.92
67	SS	2	463	1.68	0.65	–0.55	0.09	0.72	0.68
68	SS	2	463	1.55	0.64	–0.38	0.09	0.84	0.85
69	SS	2	463	1.44	0.67	0.23	0.08	0.82	0.80
70	SS	2	463	1.74	0.70	–0.57	0.09	0.59	0.50
71	SS	2	463	1.59	0.64	–0.42	0.09	0.88	0.85
72	SS	4	463	2.73	0.72	0.43	0.06	0.94	1.00
73	SS	4	463	2.82	0.74	0.39	0.06	1.06	1.13
74	SS	2	463	1.47	0.62	0.13	0.08	1.01	0.99
75	SS	2	463	1.56	0.59	–0.13	0.08	1.02	1.09
76	SS	2	463	1.47	0.53	0.15	0.08	1.18	1.21

A.4: Middle Grades

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
1	MC	1	452	0.92	0.39	–1.45	0.18	0.94	0.75
2	MC	1	452	0.90	0.47	–1.07	0.16	0.80	0.53
3	MC	1	452	0.94	0.20	–1.80	0.21	1.03	1.28
4	MC	1	452	0.64	0.44	0.60	0.11	1.06	1.12
5	MC	1	452	0.52	0.40	1.31	0.11	1.05	1.06
6	MC	1	452	0.60	0.46	0.83	0.11	1.01	1.02
7	MC	1	452	0.49	0.47	1.58	0.10	0.95	0.95
8	MC	1	452	0.64	0.47	0.94	0.11	0.95	0.91
9	MC	1	452	0.72	0.51	0.06	0.12	1.03	0.99
10	MC	1	452	0.58	0.38	1.10	0.11	1.06	1.10
11	MC	1	452	0.73	0.47	0.51	0.11	0.91	0.84
12	MC	1	452	0.46	0.39	1.66	0.10	1.01	1.11
13	MC	1	452	0.71	0.52	0.40	0.12	0.90	0.93
14	MC	1	452	0.77	0.35	0.03	0.12	1.08	1.10
15	MC	1	452	0.48	0.29	1.66	0.10	1.16	1.21
16	MC	1	452	0.93	0.34	–1.46	0.18	0.81	0.68
17	MC	1	452	0.64	0.34	0.76	0.11	1.12	1.07
18	MC	1	452	0.46	0.16	1.71	0.10	1.26	1.53
19	MC	1	452	0.67	0.34	0.59	0.11	1.11	1.16
20	MC	1	452	0.35	0.22	2.29	0.11	1.14	1.45

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
21	MC	1	452	0.95	0.19	–2.46	0.27	1.45	2.11
22	MC	1	452	0.94	0.32	–1.66	0.20	0.87	0.80
23	MC	1	452	0.72	0.45	0.14	0.12	1.05	1.04
24	MC	1	452	0.47	0.11	1.67	0.10	1.35	1.56
25	MC	1	452	0.78	0.34	–0.08	0.13	1.08	1.15
26	MC	1	452	0.54	0.00	1.29	0.11	1.49	1.61
27	MC	1	452	0.78	0.33	–0.06	0.13	1.10	1.09
28	MC	1	452	0.76	0.49	0.16	0.12	0.91	0.83
29	MC	1	452	0.79	0.39	–0.21	0.13	1.09	1.02
30	MC	1	452	0.55	0.58	1.11	0.11	0.86	0.85
31	MC	1	452	0.61	0.35	0.59	0.11	1.21	1.30
32	MC	1	452	0.65	0.50	1.00	0.11	0.89	0.86
33	MC	1	452	0.38	0.05	1.75	0.10	1.36	1.62
34	MC	1	452	0.74	0.43	0.19	0.12	1.01	0.97
35	MC	1	452	0.67	0.51	0.63	0.11	0.92	0.87
36	MC	1	452	0.56	0.58	1.20	0.11	0.83	0.81
37	MC	1	452	0.62	0.32	0.87	0.11	1.13	1.12
38	MC	1	452	0.69	0.50	0.53	0.11	0.93	0.93
39	MC	1	452	0.58	0.44	1.10	0.11	1.00	0.99
40	MC	1	452	0.67	0.47	0.64	0.11	0.97	0.97
41	MC	1	452	0.52	0.43	1.41	0.10	0.99	1.00
42	MC	1	452	0.37	0.29	2.15	0.11	1.08	1.24
43	MC	1	452	0.65	0.46	0.74	0.11	0.98	0.94
44	MC	1	452	0.70	0.33	0.45	0.11	1.12	1.12

Appendix A: Item-Level Statistics by Level and Subtest

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
45	MC	1	452	0.84	0.22	–0.48	0.14	1.09	1.47
46	MC	1	452	0.93	0.42	–1.56	0.19	0.91	0.53
47	MC	1	452	0.89	0.45	–1.04	0.16	0.93	0.65
48	MC	1	452	0.71	0.56	0.36	0.12	0.87	0.78
49	MC	1	452	0.54	0.51	1.06	0.11	0.95	0.91
50	MC	1	452	0.41	0.32	1.85	0.11	1.03	1.25
51	MC	1	452	0.61	0.52	0.94	0.11	0.91	0.91
52	MC	1	452	0.32	0.34	2.44	0.11	0.99	1.21
53	MC	1	452	0.55	0.52	1.27	0.11	0.88	0.90
54	MC	1	452	0.68	0.43	0.55	0.11	1.01	0.94
55	MC	1	452	0.61	0.49	0.90	0.11	0.96	0.93
56	MC	1	452	0.69	0.58	0.53	0.11	0.85	0.77
57	MC	1	452	0.64	0.41	0.76	0.11	1.04	1.01
58	MC	1	452	0.48	0.28	1.59	0.10	1.12	1.30
59	MC	1	452	0.44	0.14	1.80	0.11	1.28	1.52
60	MC	1	452	0.54	0.44	1.34	0.11	0.98	1.01
61	MC	1	452	0.40	0.31	1.75	0.10	1.09	1.16
62	MC	1	452	0.40	0.34	1.99	0.11	1.06	1.25
63	MC	1	452	0.53	0.40	1.35	0.11	1.04	1.04
64	MC	1	452	0.47	0.30	1.44	0.10	1.14	1.16
65	MC	1	452	0.58	0.36	1.13	0.11	1.09	1.07
66	MC	1	452	0.27	0.13	2.72	0.12	1.21	1.58
67	MC	1	452	0.37	0.14	2.14	0.11	1.24	1.58
68	MC	1	452	0.51	0.17	1.44	0.10	1.29	1.40

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
69	ER	4	452	2.65	0.74	0.62	0.06	0.87	0.87
70	ER	4	452	2.44	0.64	0.97	0.06	1.00	1.02

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
71	SS	2	452	1.77	0.54	–0.44	0.10	0.89	0.98
72	SS	2	452	1.77	0.50	–0.48	0.10	0.96	1.10
73	SS	2	452	1.46	0.68	0.71	0.08	0.93	0.84
74	SS	2	452	1.62	0.71	0.09	0.08	0.73	0.68
75	SS	2	452	1.64	0.63	0.20	0.08	0.88	0.71
76	SS	2	452	1.38	0.72	0.63	0.08	0.76	0.76
77	SS	2	452	1.42	0.77	0.54	0.08	0.66	0.60
78	SS	2	452	1.55	0.77	0.21	0.08	0.64	0.57
79	SS	2	452	1.59	0.77	0.19	0.08	0.74	0.64
80	SS	2	452	1.37	0.72	0.59	0.08	0.73	0.81
81	SS	2	452	1.52	0.76	0.26	0.08	0.81	0.75
82	SS	4	452	2.77	0.82	0.85	0.06	0.94	1.00
83	SS	4	452	2.53	0.75	0.93	0.05	0.85	0.93
84	SS	4	452	2.90	0.74	0.54	0.06	0.94	1.02

A.5: High School

Listening

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
1	MC	1	1035	0.85	0.48	–0.32	0.09	0.87	0.65
2	MC	1	1035	0.74	0.24	0.53	0.08	1.15	1.19
3	MC	1	1035	0.87	0.37	–0.92	0.11	1.33	1.28
4	MC	1	1035	0.69	0.22	0.70	0.07	1.21	1.37
5	MC	1	1035	0.42	0.34	2.16	0.07	1.05	1.11
6	MC	1	1035	0.34	0.28	2.56	0.07	1.08	1.22
7	MC	1	1035	0.52	0.31	1.38	0.07	1.16	1.21
8	MC	1	1035	0.40	0.15	1.96	0.07	1.23	1.35
9	MC	1	1035	0.22	0.16	2.95	0.08	1.02	1.17
10	MC	1	1035	0.50	0.26	1.77	0.07	1.15	1.22
11	MC	1	1035	0.53	0.31	1.62	0.07	1.11	1.13
12	MC	1	1035	0.37	0.27	2.39	0.07	1.10	1.25
13	MC	1	1035	0.50	0.48	1.80	0.07	0.93	0.93
14	MC	1	1035	0.86	0.29	–0.36	0.09	1.01	1.16
15	MC	1	1035	0.49	0.39	1.84	0.07	1.01	1.03
16	MC	1	1035	0.71	0.36	0.69	0.07	1.04	1.04
17	MC	1	1035	0.70	0.48	0.92	0.07	0.88	0.83
18	MC	1	1035	0.31	0.14	2.74	0.07	1.17	1.58
19	MC	1	1035	0.44	0.34	2.06	0.07	1.06	1.13
20	MC	1	1035	0.55	0.32	1.55	0.07	1.10	1.14

Writing Conventions

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
21	MC	1	1035	0.83	0.33	–0.09	0.09	1.03	0.99
22	MC	1	1035	0.71	0.44	0.82	0.07	0.94	0.87
23	MC	1	1035	0.94	0.30	–1.27	0.13	0.80	0.55
24	MC	1	1035	0.94	0.27	–1.46	0.14	0.95	0.88
25	MC	1	1035	0.71	0.38	0.67	0.08	1.02	1.00
26	MC	1	1035	0.87	0.32	–0.45	0.10	0.99	1.02
27	MC	1	1035	0.80	0.44	0.55	0.08	0.79	0.71
28	MC	1	1035	0.61	0.55	1.21	0.07	0.87	0.85
29	MC	1	1035	0.51	0.35	1.72	0.07	1.06	1.09
30	MC	1	1035	0.71	0.60	0.68	0.08	0.81	0.70
31	MC	1	1035	0.30	0.18	2.70	0.07	1.13	1.31
32	MC	1	1035	0.35	0.20	2.58	0.07	1.18	1.33
33	MC	1	1035	0.65	0.41	1.08	0.07	1.00	0.94
34	MC	1	1035	0.48	0.25	2.07	0.07	1.17	1.24
35	MC	1	1035	0.24	0.25	3.11	0.08	1.07	1.17
36	MC	1	1035	0.50	0.40	1.78	0.07	1.01	1.03
37	MC	1	1035	0.40	0.45	2.25	0.07	0.93	0.97
38	MC	1	1035	0.37	0.24	2.43	0.07	1.12	1.31
39	MC	1	1035	0.65	0.41	1.04	0.07	1.00	1.03
40	MC	1	1035	0.52	0.44	1.66	0.07	0.98	0.96
41	MC	1	1035	0.64	0.23	1.06	0.07	1.18	1.25
42	MC	1	1035	0.31	0.35	2.70	0.07	0.99	1.13
43	MC	1	1035	0.64	0.32	1.10	0.07	1.10	1.13
44	MC	1	1035	0.48	0.39	1.86	0.07	1.01	1.02

Reading

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
45	MC	1	1035	0.90	0.29	–0.86	0.11	0.99	0.95
46	MC	1	1035	0.79	0.46	0.20	0.08	0.93	0.82
47	MC	1	1035	0.87	0.25	–0.50	0.10	1.06	1.09
48	MC	1	1035	0.52	0.53	1.77	0.07	0.88	0.88
49	MC	1	1035	0.56	0.45	1.82	0.07	0.96	0.97
50	MC	1	1035	0.41	0.51	2.23	0.07	0.87	0.89
51	MC	1	1035	0.57	0.51	1.43	0.07	0.90	0.87
52	MC	1	1035	0.63	0.51	1.13	0.07	0.90	0.86
53	MC	1	1035	0.57	0.46	1.44	0.07	0.95	0.94
54	MC	1	1035	0.30	0.36	2.80	0.07	0.99	1.06
55	MC	1	1035	0.58	0.35	1.16	0.07	1.12	1.10
56	MC	1	1035	0.52	0.36	1.31	0.07	1.12	1.14
57	MC	1	1035	0.38	0.30	2.38	0.07	1.07	1.16
58	MC	1	1035	0.58	0.29	1.41	0.07	1.13	1.15
59	MC	1	1035	0.57	0.33	1.66	0.07	1.07	1.10
60	MC	1	1035	0.42	0.30	2.16	0.07	1.09	1.14
61	MC	1	1035	0.40	0.43	2.65	0.07	1.07	1.14
62	MC	1	1035	0.34	0.12	2.59	0.07	1.23	1.49
63	MC	1	1035	0.49	0.50	1.91	0.07	0.91	0.91
64	MC	1	1035	0.46	0.52	2.21	0.07	0.89	0.95
65	MC	1	1035	0.63	0.27	1.14	0.07	1.15	1.16
66	MC	1	1035	0.43	0.32	2.12	0.07	1.08	1.11
67	MC	1	1035	0.26	0.20	3.04	0.08	1.11	1.32
68	MC	1	1035	0.41	0.14	2.20	0.07	1.24	1.43

Writing

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
69	ER	4	1035	1.98	0.69	1.85	0.04	0.85	0.85
70	ER	4	1035	1.84	0.64	2.08	0.04	0.95	0.96

Speaking

Item #	Item Format	Max Points	N-Count	P-Value	Point Biserial	Rasch Difficulty	Standard Error of Rasch Diff	In Fit	Out Fit
71	SS	2	1035	1.47	0.66	0.64	0.05	0.78	0.78
72	SS	2	1035	1.48	0.67	0.63	0.05	0.74	0.76
73	SS	2	1035	1.18	0.67	1.45	0.05	0.86	0.85
74	SS	2	1035	1.45	0.67	0.72	0.05	0.76	0.75
75	SS	2	1035	1.39	0.72	0.51	0.05	1.07	1.00
76	SS	2	1035	1.61	0.61	0.33	0.06	0.80	0.83
77	SS	2	1035	1.39	0.72	0.95	0.05	0.71	0.70
78	SS	2	1035	1.40	0.71	0.92	0.05	0.71	0.72
79	SS	2	1035	1.40	0.66	0.67	0.06	0.91	0.90
80	SS	2	1035	1.28	0.71	1.22	0.05	0.79	0.77
81	SS	2	1035	1.32	0.71	1.03	0.05	0.74	0.74
82	SS	4	1035	2.24	0.69	1.53	0.03	1.01	1.01
83	SS	4	1035	2.70	0.77	1.21	0.03	0.83	0.88
84	SS	4	1035	2.17	0.62	1.58	0.03	1.27	1.27

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	–7.37	2.04	342	71.51
1	–5.85	1.09	395	38.05
2	–4.97	0.83	426	29.02
3	–4.37	0.73	447	25.41
4	–3.89	0.67	464	23.52
5	–3.46	0.64	479	22.51
6	–3.05	0.63	493	22.12
7	–2.65	0.64	507	22.26
8	–2.23	0.66	522	23.07
9	–1.77	0.71	538	24.82
10	–1.20	0.81	558	28.42
11	–0.35	1.07	588	37.59
12	1.15	2.04	640	71.30

Strand 3: Prereading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–6.39	2.03	376	70.91
1	–4.92	1.06	428	36.96
2	–4.10	0.79	456	27.72
3	–3.56	0.69	475	24.12
4	–3.13	0.64	491	22.40
5	–2.73	0.62	504	21.67
6	–2.35	0.62	518	21.56
7	–1.96	0.63	531	22.05
8	–1.55	0.66	546	23.21
9	–1.07	0.72	562	25.24
10	–0.48	0.83	583	29.05
11	0.41	1.09	614	38.22
12	1.94	2.05	668	71.65

Strand 4: Prewriting				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–7.53	2.04	337	71.23
1	–6.03	1.07	389	37.52
2	–5.18	0.81	419	28.42
3	–4.61	0.71	439	24.82
4	–4.15	0.66	455	22.93
5	–3.74	0.62	469	21.84
6	–3.36	0.60	482	21.14
7	–3.01	0.59	495	20.72
8	–2.66	0.59	507	20.48
9	–2.32	0.58	519	20.37
10	–1.98	0.58	531	20.41
11	–1.64	0.59	543	20.51
12	–1.29	0.59	555	20.69
13	–0.94	0.60	567	20.93
14	–0.58	0.61	580	21.25
15	–0.20	0.62	593	21.70
16	0.20	0.64	607	22.33
17	0.62	0.66	622	23.24
18	1.09	0.70	638	24.61
19	1.62	0.76	657	26.74
20	2.28	0.87	680	30.45
21	3.24	1.12	713	39.24
22	4.81	2.06	768	72.14

Strand 5: Speaking				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.08	2.00	352	69.97
1	–5.69	1.00	401	35.11
2	–4.98	0.72	426	25.31
3	–4.54	0.61	441	21.46
4	–4.20	0.56	453	19.50
5	–3.91	0.53	463	18.45
6	–3.64	0.51	473	17.89
7	–3.39	0.50	481	17.47
8	–3.14	0.49	490	17.08
9	–2.91	0.48	498	16.66
10	–2.69	0.46	506	16.21
11	–2.48	0.45	513	15.75
12	–2.28	0.44	520	15.37
13	–2.09	0.43	527	15.05
14	–1.91	0.42	533	14.81
15	–1.73	0.42	539	14.70
16	–1.55	0.42	546	14.74
17	–1.38	0.43	552	14.91
18	–1.19	0.44	558	15.26
19	–0.99	0.45	565	15.75
20	–0.78	0.47	573	16.52
21	–0.54	0.50	581	17.61
22	–0.27	0.55	591	19.18
23	0.07	0.62	602	21.63
24	0.52	0.74	618	25.87
25	1.26	1.02	644	35.74
26	2.68	2.01	694	70.35

Strand 10: Comprehension (Listening + Prereading)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.70	2.02	330	70.84
1	–6.24	1.05	382	36.65
2	–5.45	0.77	409	26.99
3	–4.95	0.65	427	22.86
4	–4.57	0.59	440	20.48
5	–4.25	0.54	451	18.90
6	–3.98	0.51	461	17.78
7	–3.73	0.49	469	16.98
8	–3.51	0.47	477	16.38
9	–3.29	0.46	485	16.00
10	–3.09	0.45	492	15.72
11	–2.89	0.44	499	15.54
12	–2.69	0.44	506	15.51
13	–2.49	0.44	513	15.54
14	–2.30	0.45	520	15.72
15	–2.09	0.46	527	16.00
16	–1.88	0.47	534	16.42
17	–1.65	0.49	542	17.01
18	–1.40	0.51	551	17.82
19	–1.13	0.54	561	18.94
20	–0.81	0.59	572	20.51
21	–0.43	0.65	585	22.89
22	0.08	0.77	603	27.06
23	0.87	1.05	630	36.72
24	2.33	2.03	682	70.91

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.94	2.01	322	70.49
1	–6.51	1.03	372	35.91
2	–5.76	0.74	398	25.94
3	–5.31	0.62	414	21.60
4	–4.98	0.55	426	19.08
5	–4.71	0.50	435	17.43
6	–4.47	0.47	443	16.31
7	–4.27	0.44	451	15.51
8	–4.08	0.43	457	14.95
9	–3.90	0.42	463	14.53
10	–3.73	0.41	469	14.21
11	–3.57	0.40	475	13.97
12	–3.41	0.39	481	13.79
13	–3.26	0.39	486	13.65
14	–3.11	0.39	491	13.51
15	–2.96	0.38	496	13.37
16	–2.82	0.38	501	13.23
17	–2.67	0.37	506	13.09
18	–2.54	0.37	511	12.95
19	–2.40	0.37	516	12.85
20	–2.27	0.36	521	12.74
21	–2.13	0.36	525	12.67
22	–2.00	0.36	530	12.64
23	–1.87	0.36	534	12.64
24	–1.74	0.36	539	12.67
25	–1.61	0.37	544	12.78
26	–1.47	0.37	548	12.92
27	–1.34	0.38	553	13.13
28	–1.19	0.38	558	13.44
29	–1.04	0.40	564	13.83
30	–0.88	0.41	569	14.35
31	–0.70	0.43	575	15.02
32	–0.51	0.45	582	15.89
33	–0.28	0.49	590	17.08
34	–0.02	0.54	599	18.76
35	0.30	0.61	611	21.28
36	0.74	0.73	626	25.59
37	1.47	1.02	652	35.60
38	2.88	2.01	701	70.28

Strand 13: Total Test				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–8.58	2.01	300	70.28
1	–7.17	1.02	349	35.60
2	–6.44	0.73	375	25.55
3	–6.00	0.61	390	21.18
4	–5.68	0.53	401	18.59
5	–5.42	0.48	410	16.84
6	–5.21	0.45	418	15.58
7	–5.02	0.42	424	14.60
8	–4.86	0.40	430	13.86
9	–4.71	0.38	435	13.23
10	–4.57	0.36	440	12.74
11	–4.44	0.35	444	12.29
12	–4.32	0.34	449	11.94
13	–4.21	0.33	453	11.66
14	–4.10	0.33	456	11.41
15	–4.00	0.32	460	11.20
16	–3.90	0.31	464	10.99
17	–3.80	0.31	467	10.85
18	–3.70	0.31	470	10.71
19	–3.61	0.30	474	10.57
20	–3.52	0.30	477	10.47
21	–3.43	0.30	480	10.40
22	–3.34	0.29	483	10.29
23	–3.26	0.29	486	10.22
24	–3.17	0.29	489	10.15
25	–3.09	0.29	492	10.08
26	–3.01	0.29	495	10.01
27	–2.92	0.29	498	9.98
28	–2.84	0.28	501	9.91
29	–2.76	0.28	503	9.87
30	–2.68	0.28	506	9.84
31	–2.60	0.28	509	9.77
32	–2.53	0.28	512	9.73
33	–2.45	0.28	514	9.73
34	–2.37	0.28	517	9.70
35	–2.30	0.28	520	9.66
36	–2.22	0.28	522	9.66
37	–2.14	0.28	525	9.66
38	–2.07	0.28	528	9.63
39	–1.99	0.28	530	9.66
40	–1.91	0.28	533	9.66

Strand 13: Total Test *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
41	–1.84	0.28	536	9.70
42	–1.76	0.28	538	9.70
43	–1.68	0.28	541	9.77
44	–1.60	0.28	544	9.80
45	–1.53	0.28	547	9.87
46	–1.45	0.28	549	9.94
47	–1.36	0.29	552	10.01
48	–1.28	0.29	555	10.12
49	–1.20	0.29	558	10.26
50	–1.11	0.30	561	10.36
51	–1.02	0.30	564	10.54
52	–0.93	0.31	568	10.71
53	–0.83	0.31	571	10.89
54	–0.73	0.32	574	11.10
55	–0.63	0.32	578	11.34
56	–0.52	0.33	582	11.62
57	–0.41	0.34	586	11.90
58	–0.29	0.35	590	12.25
59	–0.16	0.36	594	12.64
60	–0.03	0.37	599	13.09
61	0.12	0.39	604	13.58
62	0.28	0.40	610	14.14
63	0.45	0.42	616	14.84
64	0.64	0.45	622	15.61
65	0.85	0.47	630	16.56
66	1.09	0.51	638	17.71
67	1.36	0.55	648	19.18
68	1.70	0.60	659	21.11
69	2.10	0.68	674	23.80
70	2.65	0.80	693	28.14
71	3.50	1.08	722	37.73
72	5.01	2.04	775	71.47

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.53	2.04	337	71.23
1	–6.03	1.07	389	37.52
2	–5.18	0.81	419	28.42
3	–4.61	0.71	439	24.82
4	–4.15	0.66	455	22.93
5	–3.74	0.62	469	21.84
6	–3.36	0.60	482	21.14
7	–3.01	0.59	495	20.72
8	–2.66	0.59	507	20.48
9	–2.32	0.58	519	20.37
10	–1.98	0.58	531	20.41
11	–1.64	0.59	543	20.51
12	–1.29	0.59	555	20.69
13	–0.94	0.60	567	20.93
14	–0.58	0.61	580	21.25
15	–0.20	0.62	593	21.70
16	0.20	0.64	607	22.33
17	0.62	0.66	622	23.24
18	1.09	0.70	638	24.61
19	1.62	0.76	657	26.74
20	2.28	0.87	680	30.45
21	3.24	1.12	713	39.24
22	4.81	2.06	768	72.14

B.2: Primary

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–7.53	2.02	336	70.84
1	–6.07	1.05	387	36.72
2	–5.28	0.78	415	27.20
3	–4.77	0.67	433	23.28
4	–4.37	0.60	447	21.14
5	–4.02	0.57	459	19.85
6	–3.71	0.55	470	19.08
7	–3.42	0.53	480	18.66
8	–3.14	0.53	490	18.48
9	–2.86	0.53	500	18.55
10	–2.58	0.54	510	18.80
11	–2.28	0.55	520	19.25
12	–1.97	0.57	531	19.92
13	–1.63	0.59	543	20.72
14	–1.27	0.62	556	21.70
15	–0.86	0.65	570	22.86
16	–0.41	0.69	586	24.22
17	0.11	0.75	604	26.18
18	0.74	0.85	626	29.72
19	1.65	1.10	658	38.54
20	3.19	2.05	712	71.79

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–5.18	2.03	419	71.09
1	–3.70	1.06	471	37.07
2	–2.88	0.79	499	27.55
3	–2.36	0.67	517	23.45
4	–1.96	0.60	532	21.11
5	–1.62	0.56	543	19.60
6	–1.32	0.53	554	18.55
7	–1.05	0.51	563	17.78
8	–0.80	0.49	572	17.29
9	–0.56	0.48	580	16.94
10	–0.33	0.48	588	16.80
11	–0.10	0.48	596	16.77
12	0.13	0.48	605	16.94
13	0.37	0.49	613	17.29
14	0.62	0.51	622	17.89
15	0.89	0.54	631	18.80
16	1.20	0.58	642	20.20
17	1.57	0.64	655	22.44
18	2.05	0.76	672	26.50
19	2.82	1.03	699	36.19
20	4.26	2.02	749	70.60

Strand 5: Speaking				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–5.59	2.00	405	70.14
1	–4.18	1.01	454	35.32
2	–3.47	0.72	478	25.17
3	–3.05	0.59	493	20.76
4	–2.74	0.52	504	18.17
5	–2.50	0.47	513	16.42
6	–2.30	0.43	520	15.19
7	–2.12	0.41	526	14.28
8	–1.96	0.39	531	13.55
9	–1.82	0.37	536	12.99
10	–1.68	0.36	541	12.57
11	–1.56	0.35	545	12.22
12	–1.44	0.34	550	11.94
13	–1.32	0.34	554	11.76
14	–1.21	0.33	558	11.62
15	–1.10	0.33	561	11.52
16	–1.00	0.33	565	11.48
17	–0.89	0.33	569	11.52
18	–0.78	0.33	573	11.59
19	–0.67	0.34	577	11.73
20	–0.55	0.34	581	11.94
21	–0.43	0.35	585	12.18
22	–0.31	0.36	589	12.57
23	–0.18	0.37	594	13.02
24	–0.03	0.39	599	13.62
25	0.13	0.41	605	14.35
26	0.31	0.44	611	15.33
27	0.52	0.47	618	16.59
28	0.77	0.53	627	18.38
29	1.08	0.60	638	21.00
30	1.51	0.73	653	25.45
31	2.24	1.02	678	35.53
32	3.65	2.01	728	70.28

Strand 10: Comprehension (Listening + Reading)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.63	2.02	333	70.70
1	–6.18	1.04	384	36.40
2	–5.40	0.76	411	26.74
3	–4.91	0.65	428	22.65
4	–4.54	0.58	441	20.30
5	–4.23	0.54	452	18.80
6	–3.96	0.51	461	17.75
7	–3.71	0.49	470	16.98
8	–3.48	0.47	478	16.42
9	–3.27	0.46	486	15.96
10	–3.07	0.45	493	15.61
11	–2.87	0.44	500	15.37
12	–2.68	0.43	506	15.12
13	–2.50	0.43	513	14.95
14	–2.31	0.42	519	14.77
15	–2.14	0.42	525	14.63
16	–1.96	0.41	531	14.49
17	–1.79	0.41	537	14.39
18	–1.63	0.41	543	14.25
19	–1.46	0.40	549	14.14
20	–1.30	0.40	555	14.04
21	–1.14	0.40	560	13.93
22	–0.98	0.40	566	13.86
23	–0.82	0.40	571	13.83
24	–0.67	0.39	577	13.79
25	–0.51	0.39	582	13.79
26	–0.36	0.40	588	13.83
27	–0.20	0.40	593	13.90
28	–0.04	0.40	599	14.04
29	0.12	0.41	604	14.21
30	0.29	0.41	610	14.49
31	0.47	0.42	616	14.84
32	0.65	0.44	623	15.30
33	0.85	0.45	630	15.89
34	1.07	0.48	637	16.70
35	1.31	0.51	646	17.82
36	1.59	0.55	656	19.39
37	1.94	0.62	668	21.81
38	2.40	0.74	684	26.04
39	3.14	1.03	710	35.88
40	4.57	2.01	760	70.42

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–7.67	2.02	332	70.63
1	–6.23	1.04	382	36.30
2	–5.46	0.76	409	26.53
3	–4.98	0.64	426	22.33
4	–4.62	0.57	438	19.92
5	–4.32	0.52	449	18.31
6	–4.06	0.49	458	17.12
7	–3.84	0.46	466	16.24
8	–3.63	0.44	473	15.51
9	–3.44	0.43	480	14.91
10	–3.27	0.41	486	14.39
11	–3.10	0.40	491	13.93
12	–2.95	0.39	497	13.55
13	–2.80	0.38	502	13.16
14	–2.66	0.37	507	12.81
15	–2.53	0.36	511	12.50
16	–2.41	0.35	516	12.18
17	–2.29	0.34	520	11.90
18	–2.18	0.33	524	11.66
19	–2.07	0.33	528	11.41
20	–1.96	0.32	531	11.20
21	–1.86	0.32	535	11.03
22	–1.76	0.31	538	10.85
23	–1.67	0.31	542	10.71
24	–1.58	0.30	545	10.57
25	–1.49	0.30	548	10.47
26	–1.40	0.30	551	10.36
27	–1.31	0.29	554	10.29
28	–1.22	0.29	557	10.26
29	–1.14	0.29	560	10.22
30	–1.05	0.29	563	10.22
31	–0.97	0.29	566	10.22
32	–0.88	0.29	569	10.26
33	–0.79	0.30	572	10.33
34	–0.71	0.30	575	10.43
35	–0.62	0.30	578	10.54
36	–0.52	0.31	582	10.71
37	–0.43	0.31	585	10.89
38	–0.33	0.32	588	11.13
39	–0.23	0.33	592	11.41
40	–0.12	0.34	596	11.73

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
41	0.00	0.35	600	12.11
42	0.12	0.36	604	12.57
43	0.26	0.38	609	13.13
44	0.41	0.39	614	13.79
45	0.57	0.42	620	14.60
46	0.76	0.45	627	15.61
47	0.97	0.48	634	16.91
48	1.23	0.53	643	18.69
49	1.56	0.61	654	21.32
50	2.00	0.73	670	25.69
51	2.74	1.02	696	35.74
52	4.15	2.01	745	70.39

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–7.81	2.01	327	70.49
1	–6.38	1.03	377	35.98
2	–5.63	0.75	403	26.08
3	–5.17	0.62	419	21.77
4	–4.82	0.55	431	19.25
5	–4.55	0.50	441	17.57
6	–4.31	0.47	449	16.31
7	–4.11	0.44	456	15.33
8	–3.93	0.42	463	14.56
9	–3.76	0.40	468	13.90
10	–3.61	0.38	474	13.34
11	–3.47	0.37	479	12.88
12	–3.34	0.36	483	12.43
13	–3.22	0.34	487	12.04
14	–3.10	0.33	491	11.69
15	–2.99	0.33	495	11.38
16	–2.89	0.32	499	11.10
17	–2.79	0.31	502	10.82
18	–2.70	0.30	506	10.57
19	–2.61	0.30	509	10.36
20	–2.52	0.29	512	10.12
21	–2.44	0.28	515	9.94
22	–2.36	0.28	517	9.73
23	–2.28	0.27	520	9.56
24	–2.21	0.27	523	9.38
25	–2.14	0.26	525	9.24
26	–2.07	0.26	528	9.10
27	–2.00	0.26	530	8.96
28	–1.94	0.25	532	8.82
29	–1.87	0.25	534	8.72
30	–1.81	0.25	537	8.61
31	–1.75	0.24	539	8.51
32	–1.69	0.24	541	8.40
33	–1.64	0.24	543	8.33
34	–1.58	0.24	545	8.23
35	–1.53	0.23	547	8.16
36	–1.47	0.23	548	8.09
37	–1.42	0.23	550	8.05
38	–1.37	0.23	552	7.98
39	–1.31	0.23	554	7.91
40	–1.26	0.23	556	7.88

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
41	–1.21	0.22	558	7.84
42	–1.16	0.22	559	7.81
43	–1.11	0.22	561	7.78
44	–1.06	0.22	563	7.74
45	–1.01	0.22	564	7.71
46	–0.97	0.22	566	7.71
47	–0.92	0.22	568	7.67
48	–0.87	0.22	570	7.67
49	–0.82	0.22	571	7.67
50	–0.77	0.22	573	7.64
51	–0.73	0.22	575	7.64
52	–0.68	0.22	576	7.67
53	–0.63	0.22	578	7.67
54	–0.58	0.22	580	7.67
55	–0.53	0.22	581	7.71
56	–0.48	0.22	583	7.71
57	–0.44	0.22	585	7.74
58	–0.39	0.22	586	7.78
59	–0.34	0.22	588	7.81
60	–0.29	0.22	590	7.84
61	–0.24	0.23	592	7.88
62	–0.19	0.23	593	7.91
63	–0.13	0.23	595	7.95
64	–0.08	0.23	597	8.02
65	–0.03	0.23	599	8.05
66	0.02	0.23	601	8.12
67	0.08	0.23	603	8.19
68	0.13	0.24	605	8.26
69	0.19	0.24	607	8.33
70	0.25	0.24	609	8.40
71	0.31	0.24	611	8.47
72	0.36	0.25	613	8.58
73	0.43	0.25	615	8.65
74	0.49	0.25	617	8.75
75	0.55	0.25	619	8.86
76	0.62	0.26	622	8.96
77	0.68	0.26	624	9.10
78	0.75	0.26	626	9.21
79	0.82	0.27	629	9.35
80	0.89	0.27	631	9.52
81	0.97	0.28	634	9.66

Strand 13: Total Test *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
82	1.05	0.28	637	9.87
83	1.13	0.29	640	10.05
84	1.21	0.29	642	10.26
85	1.30	0.30	646	10.50
86	1.39	0.31	649	10.78
87	1.49	0.32	652	11.10
88	1.60	0.33	656	11.45
89	1.71	0.34	660	11.83
90	1.83	0.35	664	12.29
91	1.95	0.37	668	12.81
92	2.10	0.39	673	13.48
93	2.25	0.41	679	14.25
94	2.43	0.44	685	15.26
95	2.64	0.47	692	16.52
96	2.88	0.52	701	18.31
97	3.19	0.60	712	20.90
98	3.62	0.72	727	25.31
99	4.34	1.01	752	35.42
100	5.75	2.01	801	70.21

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.82	2.01	431	70.46
1	–3.39	1.03	481	35.91
2	–2.64	0.74	507	26.04
3	–2.18	0.62	524	21.81
4	–1.84	0.55	536	19.39
5	–1.56	0.51	545	17.82
6	–1.31	0.48	554	16.73
7	–1.10	0.46	562	15.96
8	–0.90	0.44	569	15.40
9	–0.71	0.43	575	14.98
10	–0.53	0.42	582	14.67
11	–0.35	0.41	588	14.46
12	–0.18	0.41	594	14.32
13	–0.02	0.41	599	14.28
14	0.15	0.41	605	14.25
15	0.32	0.41	611	14.32
16	0.48	0.41	617	14.39
17	0.66	0.42	623	14.56
18	0.83	0.42	629	14.81
19	1.01	0.43	636	15.12
20	1.21	0.44	642	15.54
21	1.41	0.46	649	16.14
22	1.63	0.48	657	16.91
23	1.88	0.51	666	17.99
24	2.17	0.56	676	19.57
25	2.52	0.63	688	21.98
26	2.99	0.75	704	26.18
27	3.74	1.03	731	36.02
28	5.17	2.01	781	70.49

B.3: Elementary

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–4.76	2.03	434	71.02
1	–3.28	1.06	485	37.00
2	–2.47	0.79	514	27.48
3	–1.95	0.67	532	23.42
4	–1.55	0.60	546	21.11
5	–1.21	0.56	558	19.60
6	–0.91	0.53	568	18.59
7	–0.64	0.51	578	17.85
8	–0.39	0.50	586	17.40
9	–0.15	0.49	595	17.08
10	0.09	0.48	603	16.94
11	0.33	0.49	611	16.98
12	0.56	0.49	620	17.15
13	0.81	0.50	628	17.50
14	1.07	0.52	637	18.10
15	1.35	0.54	647	19.04
16	1.66	0.58	658	20.44
17	2.04	0.65	671	22.68
18	2.53	0.76	689	26.71
19	3.31	1.04	716	36.33
20	4.75	2.02	766	70.67

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–4.31	2.04	449	71.40
1	–2.80	1.08	502	37.73
2	–1.94	0.81	532	28.42
3	–1.38	0.70	552	24.40
4	–0.94	0.63	567	22.02
5	–0.58	0.58	580	20.41
6	–0.26	0.55	591	19.25
7	0.03	0.53	601	18.45
8	0.30	0.51	611	17.85
9	0.56	0.50	620	17.47
10	0.80	0.49	628	17.29
11	1.05	0.49	637	17.26
12	1.29	0.50	645	17.40
13	1.54	0.51	654	17.75
14	1.81	0.52	663	18.31
15	2.10	0.55	673	19.22
16	2.42	0.59	685	20.62
17	2.80	0.65	698	22.86
18	3.30	0.77	716	26.88
19	4.08	1.04	743	36.47
20	5.53	2.02	794	70.74

Strand 5: Speaking				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.73	2.01	434	70.28
1	–3.32	1.02	484	35.53
2	–2.60	0.73	509	25.45
3	–2.17	0.60	524	21.00
4	–1.85	0.53	535	18.38
5	–1.60	0.48	544	16.63
6	–1.39	0.44	551	15.37
7	–1.21	0.41	558	14.42
8	–1.05	0.39	563	13.69
9	–0.90	0.38	568	13.13
10	–0.77	0.36	573	12.71
11	–0.64	0.35	578	12.39
12	–0.52	0.35	582	12.18
13	–0.40	0.34	586	12.01
14	–0.28	0.34	590	11.90
15	–0.16	0.34	594	11.90
16	–0.05	0.34	598	11.90
17	0.07	0.34	602	12.01
18	0.19	0.35	607	12.15
19	0.31	0.35	611	12.36
20	0.44	0.36	615	12.64
21	0.57	0.37	620	12.99
22	0.71	0.38	625	13.41
23	0.87	0.40	630	13.93
24	1.03	0.42	636	14.60
25	1.22	0.44	643	15.40
26	1.42	0.47	650	16.42
27	1.66	0.51	658	17.78
28	1.95	0.56	668	19.60
29	2.30	0.64	681	22.26
30	2.78	0.76	697	26.71
31	3.57	1.05	725	36.68
32	5.03	2.03	776	70.98

Strand 10: Comprehension (Listening + Reading)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–5.27	2.02	416	70.60
1	–3.83	1.03	466	36.19
2	–3.07	0.75	493	26.39
3	–2.59	0.63	509	22.16
4	–2.24	0.56	522	19.71
5	–1.95	0.52	532	18.06
6	–1.70	0.48	541	16.87
7	–1.48	0.46	548	15.96
8	–1.28	0.44	555	15.26
9	–1.10	0.42	562	14.67
10	–0.93	0.41	568	14.18
11	–0.77	0.39	573	13.79
12	–0.61	0.38	579	13.44
13	–0.47	0.38	584	13.16
14	–0.33	0.37	588	12.95
15	–0.19	0.36	593	12.74
16	–0.06	0.36	598	12.57
17	0.07	0.36	602	12.46
18	0.19	0.35	607	12.36
19	0.32	0.35	611	12.29
20	0.44	0.35	615	12.25
21	0.56	0.35	620	12.25
22	0.68	0.35	624	12.25
23	0.81	0.35	628	12.29
24	0.93	0.35	633	12.36
25	1.06	0.36	637	12.46
26	1.19	0.36	642	12.64
27	1.32	0.37	646	12.81
28	1.45	0.37	651	13.06
29	1.60	0.38	656	13.34
30	1.75	0.39	661	13.69
31	1.90	0.40	667	14.14
32	2.07	0.42	673	14.67
33	2.26	0.44	679	15.37
34	2.46	0.47	686	16.28
35	2.69	0.50	694	17.47
36	2.97	0.55	704	19.11
37	3.30	0.62	716	21.60
38	3.76	0.74	731	25.90
39	4.50	1.02	757	35.81
40	5.92	2.01	807	70.42

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–5.45	2.01	409	70.32
1	–4.03	1.02	459	35.63
2	–3.30	0.73	484	25.59
3	–2.86	0.61	500	21.21
4	–2.54	0.53	511	18.59
5	–2.29	0.48	520	16.80
6	–2.07	0.44	527	15.51
7	–1.89	0.41	534	14.49
8	–1.73	0.39	540	13.69
9	–1.58	0.37	545	13.02
10	–1.45	0.36	549	12.46
11	–1.33	0.34	554	12.01
12	–1.21	0.33	558	11.62
13	–1.10	0.32	561	11.27
14	–1.00	0.31	565	10.99
15	–0.91	0.31	568	10.71
16	–0.81	0.30	571	10.50
17	–0.73	0.30	575	10.33
18	–0.64	0.29	578	10.19
19	–0.56	0.29	581	10.05
20	–0.47	0.28	583	9.94
21	–0.39	0.28	586	9.87
22	–0.31	0.28	589	9.80
23	–0.24	0.28	592	9.77
24	–0.16	0.28	594	9.77
25	–0.08	0.28	597	9.77
26	0.00	0.28	600	9.77
27	0.08	0.28	603	9.80
28	0.16	0.28	605	9.84
29	0.24	0.28	608	9.91
30	0.32	0.29	611	9.98
31	0.40	0.29	614	10.08
32	0.48	0.29	617	10.19
33	0.57	0.30	620	10.33
34	0.66	0.30	623	10.50
35	0.75	0.31	626	10.68
36	0.84	0.31	630	10.89
37	0.94	0.32	633	11.10
38	1.05	0.33	637	11.38
39	1.16	0.33	640	11.66
40	1.27	0.34	644	12.01

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
41	1.39	0.35	649	12.39
42	1.52	0.37	653	12.85
43	1.66	0.38	658	13.41
44	1.82	0.40	664	14.04
45	1.99	0.42	670	14.84
46	2.18	0.45	676	15.82
47	2.40	0.49	684	17.12
48	2.66	0.54	693	18.87
49	2.99	0.61	705	21.46
50	3.44	0.74	721	25.83
51	4.18	1.02	746	35.84
52	5.61	2.01	796	70.46

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–6.11	2.01	386	70.21
1	–4.71	1.01	435	35.39
2	–3.99	0.72	460	25.31
3	–3.56	0.60	475	20.86
4	–3.25	0.52	486	18.27
5	–3.01	0.47	495	16.49
6	–2.80	0.43	502	15.19
7	–2.63	0.41	508	14.18
8	–2.47	0.38	514	13.34
9	–2.33	0.36	518	12.67
10	–2.21	0.35	523	12.11
11	–2.09	0.33	527	11.62
12	–1.99	0.32	531	11.20
13	–1.89	0.31	534	10.85
14	–1.79	0.30	537	10.50
15	–1.71	0.29	540	10.22
16	–1.62	0.28	543	9.94
17	–1.54	0.28	546	9.70
18	–1.47	0.27	549	9.49
19	–1.40	0.27	551	9.28
20	–1.33	0.26	554	9.10
21	–1.26	0.26	556	8.93
22	–1.20	0.25	558	8.79
23	–1.13	0.25	560	8.65
24	–1.07	0.24	562	8.51
25	–1.01	0.24	564	8.40
26	–0.96	0.24	566	8.30
27	–0.90	0.23	568	8.19
28	–0.85	0.23	570	8.09
29	–0.79	0.23	572	8.02
30	–0.74	0.23	574	7.95
31	–0.69	0.23	576	7.88
32	–0.64	0.22	578	7.81
33	–0.59	0.22	579	7.74
34	–0.54	0.22	581	7.71
35	–0.49	0.22	583	7.64
36	–0.45	0.22	584	7.60
37	–0.40	0.22	586	7.57
38	–0.35	0.22	588	7.53
39	–0.31	0.21	589	7.50
40	–0.26	0.21	591	7.50

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
41	–0.22	0.21	592	7.46
42	–0.17	0.21	594	7.46
43	–0.12	0.21	596	7.43
44	–0.08	0.21	597	7.43
45	–0.03	0.21	599	7.43
46	0.01	0.21	600	7.43
47	0.06	0.21	602	7.43
48	0.10	0.21	604	7.43
49	0.15	0.21	605	7.43
50	0.19	0.21	607	7.46
51	0.24	0.21	608	7.46
52	0.28	0.21	610	7.50
53	0.33	0.21	612	7.50
54	0.38	0.22	613	7.53
55	0.42	0.22	615	7.57
56	0.47	0.22	616	7.57
57	0.52	0.22	618	7.60
58	0.56	0.22	620	7.64
59	0.61	0.22	621	7.67
60	0.66	0.22	623	7.74
61	0.71	0.22	625	7.78
62	0.76	0.22	627	7.81
63	0.81	0.23	628	7.88
64	0.86	0.23	630	7.91
65	0.91	0.23	632	7.98
66	0.96	0.23	634	8.05
67	1.02	0.23	636	8.12
68	1.07	0.23	638	8.19
69	1.13	0.24	639	8.26
70	1.18	0.24	641	8.33
71	1.24	0.24	643	8.44
72	1.30	0.24	646	8.51
73	1.36	0.25	648	8.61
74	1.42	0.25	650	8.72
75	1.48	0.25	652	8.82
76	1.55	0.26	654	8.96
77	1.62	0.26	657	9.07
78	1.68	0.26	659	9.21
79	1.76	0.27	661	9.38
80	1.83	0.27	664	9.56
81	1.90	0.28	667	9.73

Strand 13: Total Test *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
82	1.98	0.28	669	9.91
83	2.07	0.29	672	10.12
84	2.15	0.30	675	10.36
85	2.24	0.30	678	10.64
86	2.34	0.31	682	10.92
87	2.44	0.32	685	11.27
88	2.54	0.33	689	11.62
89	2.66	0.34	693	12.04
90	2.78	0.36	697	12.53
91	2.92	0.37	702	13.09
92	3.06	0.39	707	13.79
93	3.23	0.42	713	14.60
94	3.42	0.45	720	15.61
95	3.63	0.48	727	16.91
96	3.89	0.53	736	18.69
97	4.21	0.61	747	21.28
98	4.66	0.73	763	25.69
99	5.39	1.02	789	35.70
100	6.81	2.01	838	70.35

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.93	2.02	427	70.77
1	–3.48	1.04	478	36.51
2	–2.70	0.76	505	26.71
3	–2.22	0.64	522	22.44
4	–1.85	0.57	535	19.92
5	–1.55	0.52	546	18.24
6	–1.30	0.49	554	17.05
7	–1.08	0.46	562	16.17
8	–0.87	0.44	569	15.54
9	–0.68	0.43	576	15.09
10	–0.50	0.42	583	14.77
11	–0.32	0.42	589	14.60
12	–0.15	0.42	595	14.53
13	0.02	0.42	601	14.53
14	0.20	0.42	607	14.63
15	0.37	0.42	613	14.77
16	0.55	0.43	619	14.98
17	0.74	0.44	626	15.30
18	0.94	0.45	633	15.65
19	1.14	0.46	640	16.14
20	1.36	0.48	648	16.73
21	1.60	0.50	656	17.47
22	1.87	0.53	665	18.48
23	2.16	0.57	676	19.81
24	2.51	0.62	688	21.63
25	2.94	0.69	703	24.26
26	3.50	0.82	723	28.53
27	4.37	1.09	753	38.01
28	5.90	2.05	806	71.61

B.4: Middle Grades

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–4.55	2.04	441	71.23
1	–3.06	1.07	493	37.49
2	–2.22	0.81	522	28.21
3	–1.66	0.70	542	24.33
4	–1.22	0.63	557	22.12
5	–0.85	0.59	570	20.62
6	–0.52	0.56	582	19.53
7	–0.22	0.54	592	18.76
8	0.06	0.52	602	18.17
9	0.32	0.51	611	17.78
10	0.58	0.50	620	17.57
11	0.83	0.50	629	17.54
12	1.08	0.50	638	17.64
13	1.34	0.51	647	17.96
14	1.61	0.53	656	18.52
15	1.90	0.55	667	19.39
16	2.23	0.59	678	20.76
17	2.62	0.66	692	22.96
18	3.12	0.77	709	26.95
19	3.91	1.04	737	36.54
20	5.36	2.02	788	70.77

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–4.08	2.04	457	71.26
1	–2.59	1.07	509	37.38
2	–1.76	0.80	539	27.83
3	–1.22	0.68	557	23.66
4	–0.81	0.61	572	21.18
5	–0.48	0.56	583	19.50
6	–0.18	0.52	594	18.31
7	0.08	0.50	603	17.40
8	0.31	0.48	611	16.73
9	0.54	0.46	619	16.24
10	0.75	0.45	626	15.89
11	0.95	0.45	633	15.68
12	1.15	0.44	640	15.54
13	1.35	0.44	647	15.54
14	1.54	0.45	654	15.61
15	1.75	0.45	661	15.86
16	1.96	0.46	668	16.17
17	2.18	0.48	676	16.70
18	2.41	0.50	684	17.43
19	2.68	0.53	694	18.45
20	2.98	0.57	704	19.95
21	3.34	0.64	717	22.30
22	3.82	0.76	734	26.43
23	4.58	1.03	760	36.19
24	6.02	2.02	811	70.60

Strand 5: Speaking				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.09	1.99	457	69.72
1	–2.72	0.99	505	34.62
2	–2.04	0.70	529	24.40
3	–1.65	0.57	542	19.99
4	–1.36	0.50	552	17.43
5	–1.14	0.45	560	15.75
6	–0.95	0.42	567	14.56
7	–0.79	0.39	572	13.65
8	–0.64	0.37	577	12.99
9	–0.51	0.36	582	12.46
10	–0.39	0.35	586	12.08
11	–0.27	0.34	590	11.73
12	–0.16	0.33	594	11.52
13	–0.06	0.32	598	11.34
14	0.05	0.32	602	11.20
15	0.15	0.32	605	11.17
16	0.25	0.32	609	11.13
17	0.35	0.32	612	11.17
18	0.46	0.32	616	11.27
19	0.56	0.33	620	11.41
20	0.67	0.33	623	11.59
21	0.78	0.34	627	11.83
22	0.90	0.35	631	12.11
23	1.02	0.36	636	12.46
24	1.16	0.37	640	12.92
25	1.30	0.38	645	13.44
26	1.45	0.40	651	14.07
27	1.62	0.43	657	14.88
28	1.82	0.45	664	15.86
29	2.04	0.49	671	17.15
30	2.30	0.54	681	18.90
31	2.63	0.61	692	21.49
32	3.08	0.74	708	25.87
33	3.82	1.02	734	35.84
34	5.25	2.01	784	70.42

Strand 10: Comprehension (Listening + Reading)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–5.06	2.02	423	70.63
1	–3.61	1.04	474	36.30
2	–2.84	0.76	501	26.57
3	–2.36	0.64	517	22.37
4	–2.00	0.57	530	19.92
5	–1.70	0.52	540	18.27
6	–1.45	0.49	549	17.08
7	–1.22	0.46	557	16.14
8	–1.02	0.44	564	15.40
9	–0.83	0.42	571	14.77
10	–0.66	0.41	577	14.25
11	–0.50	0.40	583	13.83
12	–0.35	0.38	588	13.44
13	–0.20	0.37	593	13.09
14	–0.07	0.37	598	12.81
15	0.07	0.36	602	12.57
16	0.19	0.35	607	12.36
17	0.32	0.35	611	12.18
18	0.44	0.34	615	12.04
19	0.55	0.34	619	11.90
20	0.67	0.34	623	11.83
21	0.78	0.34	627	11.76
22	0.90	0.33	631	11.69
23	1.01	0.33	635	11.69
24	1.12	0.33	639	11.69
25	1.23	0.33	643	11.69
26	1.34	0.34	647	11.76
27	1.46	0.34	651	11.83
28	1.57	0.34	655	11.94
29	1.69	0.34	659	12.04
30	1.81	0.35	663	12.22
31	1.93	0.36	668	12.43
32	2.06	0.36	672	12.67
33	2.20	0.37	677	12.99
34	2.34	0.38	682	13.37
35	2.49	0.40	687	13.83
36	2.65	0.41	693	14.39
37	2.83	0.43	699	15.09
38	3.03	0.46	706	16.00
39	3.25	0.49	714	17.22
40	3.52	0.54	723	18.90

Strand 10: Comprehension (Listening + Reading) *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
41	3.85	0.61	735	21.39
42	4.29	0.74	750	25.73
43	5.03	1.02	776	35.67
44	6.44	2.01	826	70.35

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–5.05	2.01	423	70.42
1	–3.62	1.02	473	35.77
2	–2.89	0.74	499	25.73
3	–2.44	0.61	515	21.28
4	–2.12	0.53	526	18.62
5	–1.86	0.48	535	16.80
6	–1.65	0.44	542	15.47
7	–1.47	0.41	549	14.42
8	–1.31	0.39	554	13.58
9	–1.17	0.37	559	12.88
10	–1.04	0.35	564	12.32
11	–0.92	0.34	568	11.83
12	–0.81	0.33	572	11.41
13	–0.70	0.32	575	11.06
14	–0.61	0.31	579	10.78
15	–0.51	0.30	582	10.50
16	–0.43	0.29	585	10.29
17	–0.34	0.29	588	10.12
18	–0.26	0.28	591	9.94
19	–0.18	0.28	594	9.80
20	–0.10	0.28	596	9.70
21	–0.02	0.28	599	9.63
22	0.05	0.27	602	9.56
23	0.13	0.27	604	9.49
24	0.20	0.27	607	9.45
25	0.27	0.27	610	9.45
26	0.35	0.27	612	9.45
27	0.42	0.27	615	9.49
28	0.49	0.27	617	9.52
29	0.57	0.27	620	9.56
30	0.64	0.28	622	9.63

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
31	0.72	0.28	625	9.73
32	0.80	0.28	628	9.80
33	0.88	0.28	631	9.94
34	0.96	0.29	634	10.05
35	1.04	0.29	636	10.22
36	1.13	0.30	640	10.40
37	1.22	0.30	643	10.57
38	1.31	0.31	646	10.78
39	1.41	0.32	649	11.03
40	1.51	0.32	653	11.27
41	1.62	0.33	657	11.59
42	1.73	0.34	661	11.94
43	1.85	0.35	665	12.32
44	1.98	0.37	669	12.78
45	2.12	0.38	674	13.30
46	2.27	0.40	679	13.93
47	2.44	0.42	685	14.70
48	2.63	0.45	692	15.68
49	2.84	0.48	700	16.94
50	3.10	0.53	709	18.66
51	3.42	0.61	720	21.25
52	3.86	0.73	735	25.59
53	4.59	1.02	761	35.60
54	6.01	2.01	810	70.32

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–5.85	2.01	395	70.32
1	–4.44	1.02	445	35.60
2	–3.71	0.73	470	25.52
3	–3.28	0.60	485	21.11
4	–2.96	0.53	496	18.48
5	–2.71	0.48	505	16.70
6	–2.50	0.44	513	15.37
7	–2.32	0.41	519	14.32
8	–2.16	0.39	524	13.48
9	–2.02	0.37	529	12.81
10	–1.89	0.35	534	12.22
11	–1.77	0.33	538	11.69
12	–1.67	0.32	542	11.24
13	–1.57	0.31	545	10.85
14	–1.47	0.30	548	10.50
15	–1.39	0.29	551	10.19
16	–1.30	0.28	554	9.91
17	–1.22	0.28	557	9.66
18	–1.15	0.27	560	9.42
19	–1.08	0.26	562	9.21
20	–1.01	0.26	565	9.03
21	–0.95	0.25	567	8.86
22	–0.88	0.25	569	8.68
23	–0.82	0.24	571	8.54
24	–0.76	0.24	573	8.40
25	–0.71	0.24	575	8.26
26	–0.65	0.23	577	8.16
27	–0.60	0.23	579	8.05
28	–0.55	0.23	581	7.95
29	–0.49	0.22	583	7.84
30	–0.44	0.22	584	7.74
31	–0.40	0.22	586	7.67
32	–0.35	0.22	588	7.60
33	–0.30	0.22	589	7.53
34	–0.26	0.21	591	7.46
35	–0.21	0.21	593	7.39
36	–0.17	0.21	594	7.36
37	–0.12	0.21	596	7.29
38	–0.08	0.21	597	7.25
39	–0.04	0.21	599	7.22
40	0.01	0.21	600	7.18

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
41	0.05	0.20	602	7.15
42	0.09	0.20	603	7.11
43	0.13	0.20	605	7.08
44	0.17	0.20	606	7.08
45	0.21	0.20	607	7.04
46	0.25	0.20	609	7.04
47	0.30	0.20	610	7.01
48	0.34	0.20	612	7.01
49	0.38	0.20	613	7.01
50	0.42	0.20	615	7.01
51	0.46	0.20	616	7.01
52	0.50	0.20	617	7.01
53	0.54	0.20	619	7.01
54	0.58	0.20	620	7.01
55	0.62	0.20	622	7.01
56	0.66	0.20	623	7.04
57	0.70	0.20	624	7.04
58	0.74	0.20	626	7.04
59	0.78	0.20	627	7.08
60	0.82	0.20	629	7.08
61	0.86	0.20	630	7.11
62	0.90	0.20	632	7.15
63	0.94	0.20	633	7.15
64	0.99	0.21	635	7.18
65	1.03	0.21	636	7.22
66	1.07	0.21	637	7.25
67	1.11	0.21	639	7.29
68	1.16	0.21	641	7.32
69	1.20	0.21	642	7.36
70	1.25	0.21	644	7.39
71	1.29	0.21	645	7.46
72	1.34	0.21	647	7.50
73	1.38	0.22	648	7.57
74	1.43	0.22	650	7.60
75	1.48	0.22	652	7.67
76	1.53	0.22	653	7.74
77	1.58	0.22	655	7.78
78	1.63	0.22	657	7.84
79	1.68	0.23	659	7.95
80	1.73	0.23	661	8.02
81	1.78	0.23	662	8.09

Strand 13: Total Test *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
82	1.84	0.23	664	8.19
83	1.89	0.24	666	8.26
84	1.95	0.24	668	8.37
85	2.01	0.24	670	8.47
86	2.07	0.25	672	8.61
87	2.13	0.25	675	8.72
88	2.19	0.25	677	8.86
89	2.26	0.26	679	9.00
90	2.33	0.26	681	9.17
91	2.40	0.27	684	9.35
92	2.47	0.27	686	9.52
93	2.54	0.28	689	9.73
94	2.62	0.29	692	9.98
95	2.71	0.29	695	10.22
96	2.79	0.30	698	10.50
97	2.89	0.31	701	10.82
98	2.99	0.32	705	11.17
99	3.09	0.33	708	11.59
100	3.21	0.34	712	12.04
101	3.33	0.36	717	12.60
102	3.47	0.38	721	13.27
103	3.62	0.40	727	14.07
104	3.79	0.43	733	15.05
105	3.99	0.47	740	16.38
106	4.24	0.52	748	18.13
107	4.54	0.59	759	20.76
108	4.97	0.72	774	25.20
109	5.68	1.01	799	35.32
110	7.08	2.00	848	70.14

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.85	2.04	430	71.30
1	–3.36	1.07	483	37.28
2	–2.54	0.78	511	27.44
3	–2.03	0.66	529	23.00
4	–1.65	0.58	542	20.30
5	–1.34	0.53	553	18.48
6	–1.08	0.49	562	17.15
7	–0.85	0.46	570	16.17
8	–0.65	0.44	577	15.40
9	–0.46	0.42	584	14.81
10	–0.29	0.41	590	14.35
11	–0.13	0.40	596	14.00
12	0.03	0.39	601	13.72
13	0.18	0.39	606	13.55
14	0.33	0.38	612	13.41
15	0.48	0.38	617	13.34
16	0.62	0.38	622	13.34
17	0.77	0.38	627	13.37
18	0.91	0.38	632	13.44
19	1.06	0.39	637	13.58
20	1.22	0.39	643	13.79
21	1.38	0.40	648	14.04
22	1.54	0.41	654	14.39
23	1.71	0.42	660	14.81
24	1.90	0.44	666	15.33
25	2.10	0.46	674	16.00
26	2.32	0.48	681	16.87
27	2.57	0.52	690	18.06
28	2.86	0.56	700	19.71
29	3.22	0.63	713	22.16
30	3.69	0.75	729	26.39
31	4.45	1.04	756	36.23
32	5.89	2.02	806	70.63

B.5: High School

Strand 1: Listening				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–3.56	2.03	475	71.19
1	–2.07	1.07	527	37.31
2	–1.24	0.80	556	27.90
3	–0.70	0.68	575	23.94
4	–0.28	0.62	590	21.63
5	0.07	0.58	603	20.16
6	0.39	0.55	614	19.11
7	0.68	0.53	624	18.38
8	0.94	0.51	633	17.85
9	1.20	0.50	642	17.54
10	1.45	0.50	651	17.36
11	1.69	0.50	659	17.33
12	1.94	0.50	668	17.47
13	2.19	0.51	677	17.82
14	2.46	0.53	686	18.38
15	2.75	0.55	696	19.25
16	3.07	0.59	708	20.65
17	3.46	0.65	721	22.86
18	3.95	0.77	738	26.85
19	4.74	1.04	766	36.47
20	6.19	2.02	816	70.70

Strand 3: Reading				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–3.47	2.03	478	71.19
1	–1.99	1.06	531	37.24
2	–1.16	0.79	559	27.65
3	–0.64	0.67	578	23.49
4	–0.24	0.60	592	21.00
5	0.09	0.55	603	19.32
6	0.38	0.52	613	18.13
7	0.64	0.49	622	17.26
8	0.87	0.47	630	16.59
9	1.09	0.46	638	16.10
10	1.30	0.45	645	15.79
11	1.50	0.44	652	15.54
12	1.69	0.44	659	15.44
13	1.89	0.44	666	15.44
14	2.08	0.44	673	15.51
15	2.28	0.45	680	15.72
16	2.49	0.46	687	16.07
17	2.71	0.47	695	16.56
18	2.94	0.49	703	17.29
19	3.20	0.52	712	18.31
20	3.49	0.57	722	19.85
21	3.85	0.63	735	22.19
22	4.33	0.75	751	26.32
23	5.08	1.03	778	36.09
24	6.52	2.02	828	70.53

Strand 5: Speaking				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–3.71	2.01	470	70.32
1	–2.30	1.02	520	35.53
2	–1.57	0.73	545	25.38
3	–1.14	0.60	560	20.93
4	–0.83	0.52	571	18.27
5	–0.59	0.47	580	16.52
6	–0.38	0.44	587	15.23
7	–0.20	0.41	593	14.25
8	–0.05	0.39	598	13.51
9	0.10	0.37	603	12.95
10	0.23	0.36	608	12.50
11	0.35	0.35	612	12.15
12	0.47	0.34	616	11.87
13	0.58	0.33	620	11.66
14	0.69	0.33	624	11.55
15	0.80	0.33	628	11.48
16	0.91	0.33	632	11.45
17	1.02	0.33	636	11.48
18	1.13	0.33	639	11.55
19	1.24	0.33	643	11.66
20	1.35	0.34	647	11.83
21	1.47	0.34	651	12.01
22	1.59	0.35	656	12.25
23	1.71	0.36	660	12.57
24	1.84	0.37	665	12.92
25	1.99	0.38	669	13.37
26	2.14	0.40	675	13.90
27	2.30	0.42	681	14.60
28	2.49	0.44	687	15.51
29	2.70	0.48	695	16.73
30	2.95	0.53	703	18.45
31	3.27	0.60	714	21.00
32	3.70	0.73	729	25.41
33	4.42	1.01	755	35.46
34	5.83	2.01	804	70.25

Strand 10: Comprehension (Listening + Reading)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.23	2.02	452	70.60
1	–2.79	1.03	502	36.16
2	–2.03	0.75	529	26.36
3	–1.56	0.63	545	22.12
4	–1.20	0.56	558	19.64
5	–0.92	0.51	568	17.99
6	–0.67	0.48	577	16.77
7	–0.45	0.45	584	15.82
8	–0.26	0.43	591	15.09
9	–0.08	0.41	597	14.46
10	0.09	0.40	603	13.97
11	0.24	0.39	608	13.51
12	0.38	0.38	613	13.16
13	0.52	0.37	618	12.85
14	0.65	0.36	623	12.57
15	0.78	0.35	627	12.36
16	0.90	0.35	632	12.15
17	1.02	0.34	636	11.97
18	1.14	0.34	640	11.83
19	1.25	0.34	644	11.73
20	1.36	0.33	648	11.66
21	1.48	0.33	652	11.59
22	1.58	0.33	655	11.55
23	1.69	0.33	659	11.52
24	1.80	0.33	663	11.52
25	1.91	0.33	667	11.55
26	2.02	0.33	671	11.62
27	2.13	0.33	675	11.69
28	2.24	0.34	679	11.80
29	2.36	0.34	683	11.94
30	2.48	0.35	687	12.11
31	2.60	0.35	691	12.29
32	2.73	0.36	695	12.57
33	2.86	0.37	700	12.88
34	3.00	0.38	705	13.23
35	3.14	0.39	710	13.69
36	3.30	0.41	716	14.28
37	3.48	0.43	722	14.98
38	3.67	0.45	729	15.89
39	3.89	0.49	736	17.08
40	4.16	0.54	745	18.76

Strand 10: Comprehension (Listening + Reading) *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
41	4.48	0.61	757	21.28
42	4.92	0.73	772	25.62
43	5.65	1.02	798	35.60
44	7.07	2.01	847	70.28

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.34	2.01	448	70.35
1	–2.93	1.02	498	35.67
2	–2.20	0.73	523	25.62
3	–1.76	0.61	539	21.21
4	–1.43	0.53	550	18.59
5	–1.18	0.48	559	16.80
6	–0.97	0.44	566	15.47
7	–0.79	0.41	573	14.46
8	–0.62	0.39	578	13.65
9	–0.48	0.37	583	12.95
10	–0.35	0.35	588	12.39
11	–0.23	0.34	592	11.94
12	–0.11	0.33	596	11.52
13	–0.01	0.32	600	11.17
14	0.09	0.31	603	10.89
15	0.19	0.30	606	10.64
16	0.28	0.30	610	10.43
17	0.36	0.29	613	10.22
18	0.45	0.29	616	10.08
19	0.53	0.28	619	9.94
20	0.61	0.28	621	9.84
21	0.69	0.28	624	9.77
22	0.77	0.28	627	9.70
23	0.84	0.28	630	9.66
24	0.92	0.28	632	9.66
25	1.00	0.28	635	9.63
26	1.07	0.28	638	9.66
27	1.15	0.28	640	9.66
28	1.22	0.28	643	9.70
29	1.30	0.28	646	9.73
30	1.38	0.28	648	9.80

Strand 11: Oral (Listening + Speaking)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
31	1.46	0.28	651	9.87
32	1.54	0.28	654	9.94
33	1.62	0.29	657	10.05
34	1.71	0.29	660	10.15
35	1.79	0.29	663	10.29
36	1.88	0.30	666	10.43
37	1.97	0.30	669	10.61
38	2.06	0.31	672	10.78
39	2.16	0.31	676	10.99
40	2.26	0.32	679	11.24
41	2.37	0.33	683	11.52
42	2.48	0.34	687	11.83
43	2.60	0.35	691	12.22
44	2.72	0.36	695	12.67
45	2.86	0.38	700	13.20
46	3.01	0.40	705	13.83
47	3.17	0.42	711	14.60
48	3.36	0.45	718	15.58
49	3.57	0.48	725	16.84
50	3.83	0.53	734	18.59
51	4.15	0.60	745	21.14
52	4.58	0.73	760	25.55
53	5.31	1.02	786	35.56
54	6.72	2.01	835	70.28

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
0	–5.17	2.01	419	70.25
1	–3.76	1.02	468	35.53
2	–3.04	0.73	494	25.45
3	–2.61	0.60	509	21.04
4	–2.29	0.53	520	18.41
5	–2.04	0.48	529	16.63
6	–1.83	0.44	536	15.33
7	–1.65	0.41	542	14.32
8	–1.50	0.39	548	13.48
9	–1.36	0.37	553	12.81
10	–1.23	0.35	557	12.22
11	–1.11	0.34	561	11.73
12	–1.00	0.32	565	11.31
13	–0.90	0.31	568	10.92
14	–0.81	0.30	572	10.57
15	–0.72	0.29	575	10.26
16	–0.63	0.29	578	10.01
17	–0.55	0.28	581	9.73
18	–0.48	0.27	583	9.52
19	–0.40	0.27	586	9.31
20	–0.33	0.26	588	9.14
21	–0.27	0.26	591	8.93
22	–0.20	0.25	593	8.79
23	–0.14	0.25	595	8.65
24	–0.08	0.24	597	8.51
25	–0.02	0.24	599	8.37
26	0.03	0.24	601	8.23
27	0.09	0.23	603	8.12
28	0.14	0.23	605	8.02
29	0.19	0.23	607	7.95
30	0.24	0.22	609	7.84
31	0.29	0.22	610	7.78
32	0.34	0.22	612	7.71
33	0.39	0.22	614	7.64
34	0.44	0.22	615	7.57
35	0.49	0.21	617	7.50
36	0.53	0.21	619	7.46
37	0.58	0.21	620	7.39
38	0.62	0.21	622	7.36
39	0.66	0.21	623	7.32
40	0.71	0.21	625	7.29

Strand 13: Total Test				
Raw Score	Measure	Raw Score Standard Error	Scale Score	Scale Score Standard Error
41	0.75	0.21	626	7.25
42	0.79	0.21	628	7.22
43	0.84	0.21	629	7.22
44	0.88	0.21	631	7.18
45	0.92	0.21	632	7.18
46	0.96	0.20	634	7.15
47	1.01	0.20	635	7.15
48	1.05	0.20	637	7.11
49	1.09	0.20	638	7.11
50	1.13	0.20	640	7.11
51	1.17	0.20	641	7.11
52	1.21	0.20	642	7.11
53	1.25	0.20	644	7.11
54	1.30	0.20	645	7.11
55	1.34	0.20	647	7.11
56	1.38	0.20	648	7.15
57	1.42	0.20	650	7.15
58	1.46	0.20	651	7.15
59	1.50	0.21	653	7.18
60	1.55	0.21	654	7.18
61	1.59	0.21	656	7.18
62	1.63	0.21	657	7.22
63	1.67	0.21	659	7.25
64	1.72	0.21	660	7.25
65	1.76	0.21	662	7.29
66	1.80	0.21	663	7.32
67	1.85	0.21	665	7.36
68	1.89	0.21	666	7.39
69	1.94	0.21	668	7.43
70	1.98	0.21	669	7.46
71	2.03	0.21	671	7.50
72	2.07	0.22	673	7.53
73	2.12	0.22	674	7.60
74	2.17	0.22	676	7.64
75	2.22	0.22	678	7.71
76	2.27	0.22	679	7.74
77	2.31	0.22	681	7.81
78	2.37	0.23	683	7.88
79	2.42	0.23	685	7.95
80	2.47	0.23	686	8.02
81	2.52	0.23	688	8.09

Strand 13: Total Test *(continued)*				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
82	2.58	0.23	690	8.19
83	2.63	0.24	692	8.26
84	2.69	0.24	694	8.37
85	2.75	0.24	696	8.47
86	2.81	0.25	698	8.58
87	2.87	0.25	700	8.72
88	2.93	0.25	703	8.86
89	2.99	0.26	705	9.00
90	3.06	0.26	707	9.14
91	3.13	0.27	710	9.31
92	3.20	0.27	712	9.49
93	3.28	0.28	715	9.70
94	3.36	0.28	718	9.91
95	3.44	0.29	720	10.19
96	3.53	0.30	723	10.43
97	3.62	0.31	727	10.75
98	3.72	0.32	730	11.13
99	3.82	0.33	734	11.52
100	3.94	0.34	738	11.97
101	4.06	0.36	742	12.53
102	4.19	0.38	747	13.20
103	4.34	0.40	752	14.00
104	4.52	0.43	758	14.98
105	4.72	0.47	765	16.28
106	4.96	0.52	773	18.06
107	5.26	0.59	784	20.69
108	5.68	0.72	799	25.13
109	6.39	1.01	824	35.28
110	7.79	2.00	873	70.14

Strand 14: Total Writing (Writing Conventions + Writing)				
Raw Score	**Measure**	**Raw Score Standard Error**	**Scale Score**	**Scale Score Standard Error**
0	–4.17	2.03	454	71.12
1	–2.68	1.06	506	37.10
2	–1.87	0.78	535	27.44
3	–1.35	0.66	553	23.21
4	–0.96	0.59	566	20.62
5	–0.65	0.54	577	18.87
6	–0.38	0.50	587	17.57
7	–0.14	0.47	595	16.59
8	0.08	0.45	603	15.82
9	0.27	0.44	610	15.23
10	0.46	0.42	616	14.81
11	0.63	0.41	622	14.46
12	0.80	0.41	628	14.21
13	0.96	0.40	634	14.04
14	1.12	0.40	639	13.93
15	1.28	0.40	645	13.90
16	1.44	0.40	650	13.90
17	1.60	0.40	656	13.97
18	1.76	0.40	662	14.04
19	1.92	0.41	667	14.18
20	2.09	0.41	673	14.35
21	2.26	0.42	679	14.60
22	2.44	0.43	685	14.88
23	2.62	0.44	692	15.23
24	2.82	0.45	699	15.68
25	3.03	0.47	706	16.28
26	3.25	0.49	714	17.05
27	3.50	0.52	723	18.10
28	3.79	0.56	733	19.64
29	4.15	0.63	745	22.02
30	4.61	0.75	761	26.18
31	5.36	1.03	788	35.98
32	6.79	2.01	838	70.46

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 – Noon	Individual Groups Review and Discuss Threshold Descriptors
	— Sedona 1, 2, 3, 4 and Phoenix 1 rooms
Noon – 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*
Noon – 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 clues.

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 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 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 steps) 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 has 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 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 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 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 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 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 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 and 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 and 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 IV — 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, summarizing events, and reporting 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: 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												
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												
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												
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												
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												
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

Appendix D: References

Andrich, A., & Luo, G. (2004). *Modern measurement and analysis in social science*. Perth, Western Australia: Murdoch University.

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.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Publishers.

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.

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*, 179–1987.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

Mitzel, H.C., Lewis, D.M., & Green, D.R. (2001). The bookmark procedure. In G.J. Cizek (Ed.), *Setting performance standards: concepts, methods, and perspectives* (pp. 249–282). Mahwah, New Jersey. Lawrence Earlbaum Associates.

Nitko, A. J. (2004). *Educational Assessment of Students* (4th Ed.), Upper Saddle River, NJ: Person Education Inc.

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.

Thissen, D., & Steinberg, L. (1983). *A response model for multiple choice items* (Psychometric Tech. Rep. No.1). Chicago, IL: National Opinion Research Center.

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.