

TR 2008 - 12

**The Ohio Test of English Language
Acquisition (OTELA),**

Technical Manual

American Institutes for Research

August 6, 2008

Revision	Description	Revised by	Approval	Date
Int. Rel.	Edited by T. Moore	TWM		8.6.08

TR 2008 – 12

THE OHIO TEST OF ENGLISH LANGUAGE ACQUISITION (OTELA)

TECHNICAL MANUAL

MAY 8, 2006

Table of Contents

Introduction	1
1. Background.....	1
Development of the English Language Development Assessment (ELDA).....	1
ELDA Field Test and Operational Test Calibration and Equating	3
ELDA Performance Standards.....	5
Need for a Short Version of the ELDA	9
2. OTELA Test Development Framework.....	11
3. OTELA Test Characteristics.....	13
References	22

Appendices

Appendix A. Theta to ELDA Scaled Score TransformationsA-1
Appendix B. Raw Score Frequency Distributions for the 2005 ELDA Operational Tests.....B-1
Appendix C. OTELA Conversion Tables C-1

Introduction

This manual describes the development of the Ohio Test of English Language Acquisition (OTELA) and documents test characteristics of the OTELA. The first section of this manual provides background information on the development of the English Language Development Assessment (ELDA) from which the OTELA was constructed and describes the goals underlying the OTELA test development process. Section 2 describes the test specifications and form construction process for the OTELA. Finally, Section 3 presents the test characteristics and statistical properties of the OTELA.

1. Background

Development of the English Language Development Assessment (ELDA)

SCASS Assessment Goals. One of the requirements of the *No Child Left Behind Act of 2001* (NCLB) concerns the annual assessment of limited English proficient (LEP) students regarding their acquisition of and progress in developing English language proficiency in Listening, Speaking, Reading, and Writing. The English Language Development Assessment (ELDA) was designed to allow states to meet this federal requirement. The American Institutes for Research (AIR) developed the ELDA under the direction of a consortium of 18 member states of the LEP State Collaborative on Assessment and Student Standards (the LEP-SCASS) and the Council of Chief State School Officers (CCSSO).

The ELDA is a battery of tests designed to allow schools to measure annual progress in the acquisition of English language proficiency skills among non-native English-speaking students in grades 3–12. The battery consists of separate tests for Listening, Speaking, Reading, and Writing, at each of three grade clusters: 3–5, 6–8 and 9–12. The tests are aligned with state English language proficiency (ELP) standards of project member states and are constructed to provide content coverage across three academic topic areas (English Language Arts—ELA; Mathematics, Science and Technology—MST; and Social Studies—SS) and one nonacademic topic area related to the school environment (School-Environmental—S-E, which includes topics such as extracurricular activities, student health, homework, classroom management, and lunchtime, among others). Although the ELDA tests measure language skills with content drawn from age-appropriate school curricular and non-curricular sources, they are not tests of academic content. Students do not need any external or prior content-related knowledge to respond to test questions.

Because LEP students demonstrate a wide range of English language proficiencies, from an inability to communicate in English to mastery of the English language, the ELDA attempts to measure a correspondingly wide range of proficiencies. To measure these proficiencies, the ELDA includes many items, resulting in a battery of relatively long tests. In addition, the wide range of proficiencies measured necessitates imprecise measurement at any point on the scale.

Test Specifications. The ELDA was aligned with the state ELP standards through an analysis of the ELP standards of consortium states available to the project at the outset. AIR constructed, and the LEP-SCASS approved, a set of core ELP standards, which formed the basis for item design. AIR also developed a set of *benchmarks*, or specific statements of what students should know and be able to do—measures of progress toward meeting a standard—for each standard. Benchmarks were not developed for the Speaking standards. The standards and benchmarks were then used to establish test specifications, to classify items and to map items on test forms.

The content that forms the basis for test items in the ELDA is distributed across four topic areas. Approximately 25% of the items use language from each of the three curriculum domains of mathematics, science and technology; English language arts; and social studies. The remaining 25% of the items use the social language of interaction between students, teachers, other school personnel, and parents related to school issues. To represent a range of difficulties, from pre-functional to full English proficiency, test specifications called for test forms to comprise a roughly equivalent number of low-, mid- and high-difficulty items. The ELDA also strives for an approximately equal distribution of items across benchmarks.

The LEP-SCASS approved test item specifications for each of the four skill domain areas. The Listening tests consist of only multiple-choice items. Students listen to five types of texts—short phrases, short dialogues, extended dialogues, short presentations, and extended presentations—read by a narrator and actors. Extended presentations are not, however, part of the grade 3–5 Listening tests because they are developmentally inappropriate. After listening to each presentation, students respond to items assessing their comprehension of the presentation.

The Reading tests also entirely comprise multiple-choice items. Students read three types of text (short, early reading comprehension passages in cloze item format¹; instructions; and long passages) and answer comprehension questions. The Writing assessments include both multiple-choice and short and extended constructed-response items. Three broad standards—editing, revising, and planning and organizing—are assessed through multiple-choice items attached to short “student written” passages. Speaking consists only of constructed-response items. Items come in sets of four prompts, each eliciting a different speaking function: connect, tell, expand, reason.

Test Development. To develop items measuring the academic standards required by the test specifications, AIR brought together a highly competent pool of item writers, using a mix of external item writers, NAEP foreign language item writers and other internal content experts. After writers drafted and reviewed the items, the items were entered into the review protocol as part of AIR’s Item Tracking System database. AIR content experts and editors conducted a series of reviews to ensure that items matched the standards and represented relatively pure measures of the intended constructs. Items passing those reviews were brought to LEP-SCASS meetings for review, comments, revision, and approval. Items surviving this final review process were entered into the field-test item pool, and field-test forms were constructed. Following the first field-test administration, AIR performed classical item analyses and differential item functioning (DIF) analysis to evaluate the performance of field-test items. In addition, AIR used item

¹ Cloze format is a sentence where a word or some words have been removed and the examinee is asked to supply the missing word or words.

response theory (IRT) procedures to calibrate the field-test items and place them on a common scale.

Following item data review and approval of the operational item pool by the LEP-SCASS, AIR content specialists worked with AIR psychometricians to assemble operational forms based on test specifications and statistical requirements of the ELDA program. In the following sections, we briefly describe the psychometric analyses performed on the ELDA field-test and operational-test data.

ELDA Field-Test and Operational-Test Calibration and Equating

Overview. In 2004, two ELDA field-test forms for each subject and grade cluster were developed, resulting in a total of 24 field-test forms. The first ELDA operational form (Form 1) was then developed from the 2004 ELDA field test and administered in 2005. In addition to the operational assessment, a pool of new ELDA items was field tested in the spring of 2005. Three field-test forms were developed for the Reading, Listening and Speaking tests, and four field-test forms were developed for the Writing tests. The 2005 field-test items, along with a small number of unused items in the 2004 field-test item pool, were used to assemble ELDA operational Form 2 and Form 3.

Before analyzing the field-test data, AIR staff performed a number of quality control procedures to ensure the accuracy of the data. AIR conducted initial evaluations of item quality and examined potential bias in item performance, using classical item analysis procedures and analysis of differential item functioning (DIF). AIR conducted DIF analyses on all items in the field test to detect potential item bias across gender groups and different language background groups. Items flagged for out-of-range classical item statistics and DIF analyses were evaluated by the item review committee to determine whether flagged items should be included in the operational item pool.

AIR used Winsteps, publicly available IRT software from Mesa Press, to estimate ELDA item parameters. For multiple-choice items, we used the Rasch model (Rasch, 1960), and for constructed-response items, we used Masters' (1982) partial-credit model. To evaluate item fit, we examined the Infit and Outfit statistics provided by Winsteps. Misfitting items were evaluated in conjunction with the classical item analysis results to determine whether items should be included in the operational pool. Below, we describe the IRT calibration and equating for the 2004 field-test items, the 2005 operational test items and the 2005 field-test items.

2004 Field-Test Calibration. For the initial field test of items for each subject and grade cluster, the field-test administration used a randomly equivalent groups design to ensure the comparability of samples across field-test forms. Reading and Writing forms were spiraled within classrooms, whereas Speaking and Listening forms were spiraled across classrooms within schools to ensure that responses were gathered from randomly equivalent groups. To implement the randomly equivalent groups design, we conducted the IRT calibrations for each field test by setting the mean population ability to zero for each form in each grade cluster.

In addition to forms equating, the ELDA built in common items across grade clusters to facilitate vertical linking, that is, to enable the construction of a common score scale across grades.

Through the IRT calibration of common items embedded in the test forms of the adjacent grade clusters, it was possible to link the measurement scale of the three grade clusters into one scale. This was achieved by finding the transformation constants of adjacent grade clusters by using the mean-mean equating method described in Kolen and Brennan (1995). For our purpose, the grade cluster 3–5 was used as the base scale group, and the vertical scale was created by shifting the separately calibrated scales of the other two clusters by the appropriate linking constants.

To ensure the quality of the linking item pool, we used a stepwise deletion procedure when computing the linking constant. At each step, we eliminated the potential anchor item with the greatest difference between its linked and reference item parameter estimates, provided the difference was greater than the tolerance level (set to 0.3 in this study). We repeated this process until all remaining anchor items met the tolerance-level specifications. The stepwise procedure was not used to link the grade 3–5 and grade 6–8 clusters in Writing because there were only five common items between these two grade clusters. All five common Writing items were used to link these two grade clusters.

2005 Operational Test Calibration. Following the ELDA Technical Advisory Committee’s recommendation, ELDA operational Form 1 was recalibrated after the 2005 administration and serves as the reference scales. All subsequent ELDA test forms are linked to these scales.

Because the first set of operational forms were constructed to include a set of common items between adjacent grade clusters, the grade 3–5, grade 6–8 and grade 9–12 forms were jointly calibrated in a single Winsteps run for each subject, resulting in a common, vertically linked scale across grade clusters for each subject. For each Winsteps run, the mean of the item difficulty parameters was fixed to zero so that the average difficulty (i.e., average item step value) for all items across grade clusters was equal to zero within each subject for the first operational form.

2005 Field Test Calibration. For the 2005 field test, a common item design was used to allow common item equating across field-test forms and the first operational form. Following the common item design of the field test, items were jointly calibrated in a single Winsteps run for each subject and grade-cluster combination. Because all 2005 ELDA field-test forms share common items with operational Form 1, we could equate the field-test items to the operational Form 1 scale by using the common item equating method. For each field-test form within each grade cluster, shared items were fixed to their operational Form 1 scale values, while the remaining items were freely estimated. This placed all the field-test items on the operational Form 1 scale. Basically, item calibration and form linking were conducted simultaneously.

Equating Remaining 2004 Field-Test Items to the 2005 Operational Scale. A small subset of items was field tested in 2004 but was not included in the 2005 operational forms. These items were also placed on the 2005 operational ELDA scale. Because all items in the 2005 operational test came from the 2004 field-test item pool, we were able to use these items as linking items. The mean-mean procedure was used to find the linking constant. In the linking process, it is important to use a set of anchor items (i.e., common items) that are free of item parameter drift so that the linking constant can be estimated correctly.

AIR used a stepwise deletion procedure to select anchor items and calculate the linking constant needed to bring the field test items onto the reference scale defined by the first operational administration. Following this procedure, we first calculated the linking constant, using all anchor items, and then applied the linking constant to bring the items back to the reference scale. We then examined the parameter estimates of the anchor items to determine whether the difference between any adjusted or linked parameter estimates and the reference scale parameter estimates was greater than 0.3 logits. At each step, we eliminated from the anchor set the item with the greatest difference between its linked and reference item parameter estimates, provided the difference was greater than 0.3. We then computed and applied a new linking constant to the test items and again examined the parameter estimates for the remaining anchor items to determine whether any exceeded the 0.3 tolerance level. We repeated this process until all remaining anchor items met the tolerance-level specification of 0.3 logits. The linking constant was computed on the basis of this final anchor item set and then applied to the 2004 ELDA field-test item parameters.

ELDA Performance Standards

AIR developed performance level descriptors (PLDs) for each of the four language skills tested in the ELDA battery—Listening, Speaking, Reading, and Writing. At the outset of the project, the LEP-SCASS Steering Committee determined that the PLDs should contain five levels of proficiency, as indicated in Table 1. The levels range from Full English Proficiency, a level at which an LEP student is deemed to be able to function effectively and consistently through the medium of academic English in the school system (and thus ceases to be defined as LEP), to Pre-functional, a level at which an LEP student is consistently unable to communicate with any success in the English of the school environment, although the student may have some limited knowledge of English.

Table 1. ELDA Performance Levels

Level	Label
5	Full English Proficiency
4	Advanced
3	Intermediate
2	Beginners
1	Pre-functional

After the 2005 ELDA operational Form 1 was calibrated, AIR produced raw score to theta score conversions and ordered item book maps for each subject and grade cluster, which were delivered to Measurement Incorporated (MI). These files provided the basis for developing scaled scores and for standard setting. MI conducted the standard setting for the ELDA tests. The ability parameters associated with the cut scores are shown in Table 2. Appendix A lists the theta to ELDA scaled score transformation formulas.

Table 2. Ability Measures at Cut Scores for ELDA

Subject	Grade Cluster	Theta at Cut Score			
		Level 2	Level 3	Level 4	Level 5
Reading	3–5	-0.62	0.45	1.00	2.00
	6–8	-0.54	0.71	1.35	2.49
	9–12	0.16	0.86	1.58	2.66
Listening	3–5	-0.83	-0.04	0.81	1.48
	6–8	0.04	0.65	1.42	2.16
	9–12	0.06	0.70	1.51	2.53
Speaking	3–5	-1.39	-0.33	0.98	2.51
	6–8	-1.30	0.36	1.53	2.69
	9–12	-0.08	0.78	2.04	2.96
Writing	3–5	-1.14	0.08	0.97	4.74
	6–8	-0.15	0.81	1.48	3.16
	9–12	-0.57	0.60	1.45	2.71

Figures 1-4 illustrate the articulation of ELDA performance standards across grade clusters for each of the four subject areas. As the figures indicate, the articulation of cut scores across grade levels is uneven across subject areas. In Listening, for example, performance standards for students in the 6-8 grade cluster are similar to performance standards for students in the grade 9-12 grade cluster for three of the performance level cut scores. In writing, the performance standards for the 6-8 and 9-12 grade clusters are actually inverted, with performance standards for students in the grade 6-8 cluster placed uniformly higher than for students in the grade 9-12 cluster. Moreover, the Full English Proficiency (Level 5) cut for students in the grade 3-5 cluster is higher than for students in either the 6-8 or 9-12 grade cluster, indicating that students in the grade 3-5 cluster must outperform all higher grade level students to achieve Full English Proficiency in writing.

Figure 1. ELDA Cut Scores in Theta Metric – Reading

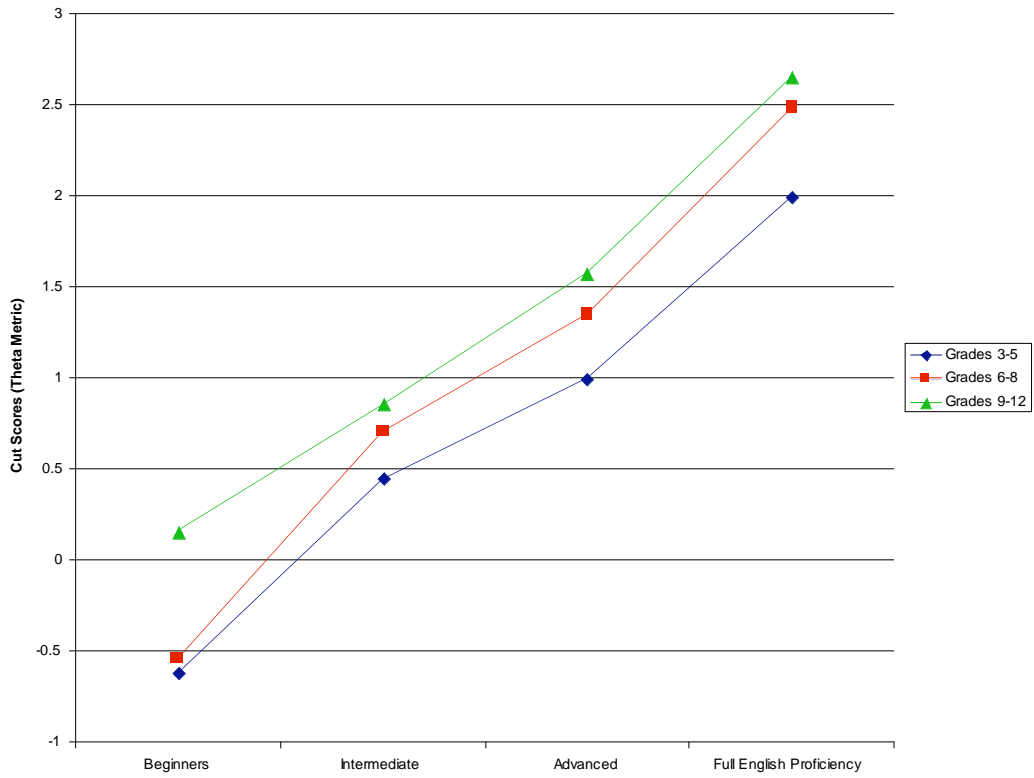


Figure 2. ELDA Cut Scores in Theta Metric – Listening

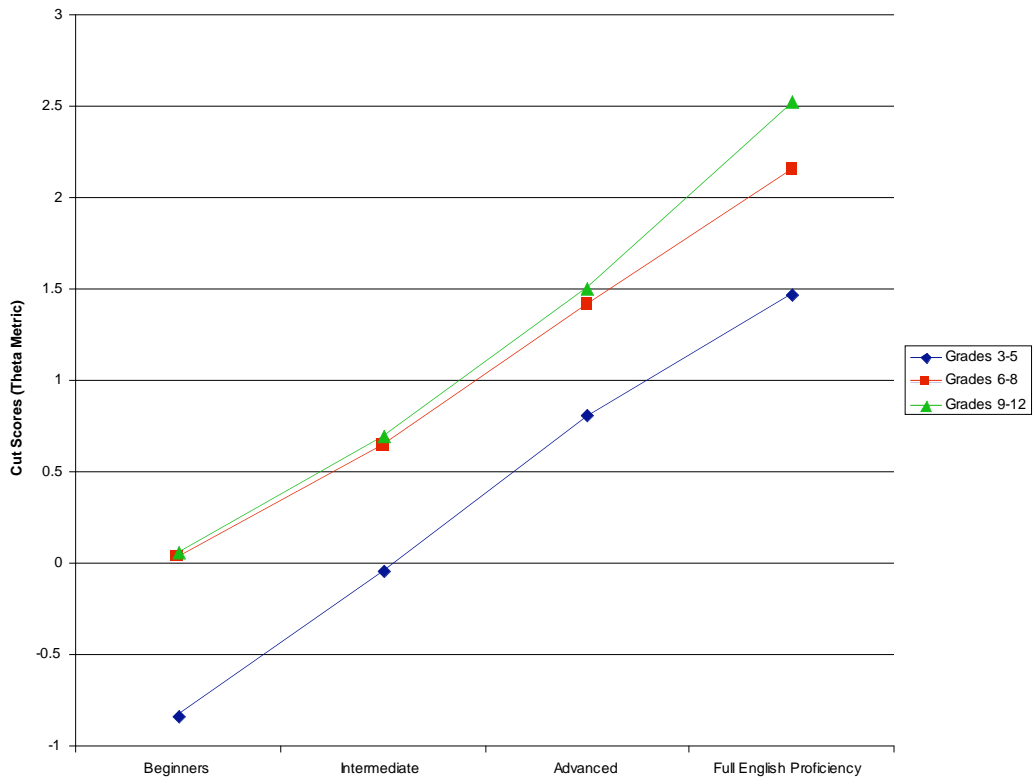


Figure 3. ELDA Cut Scores in Theta Metric – Speaking

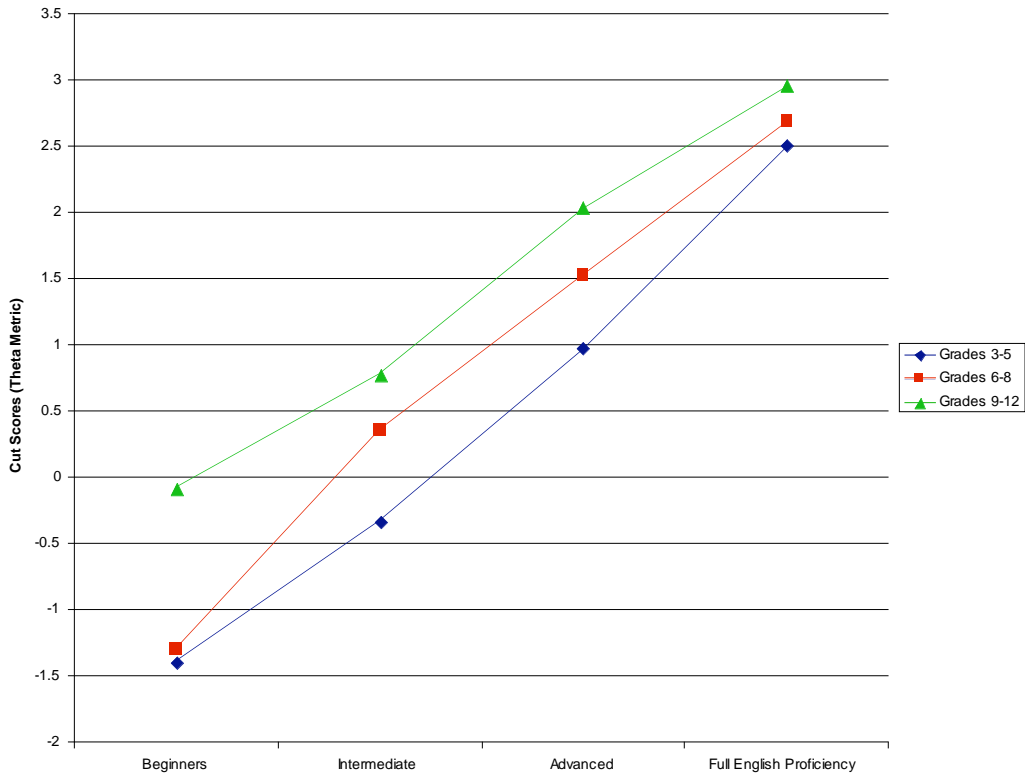
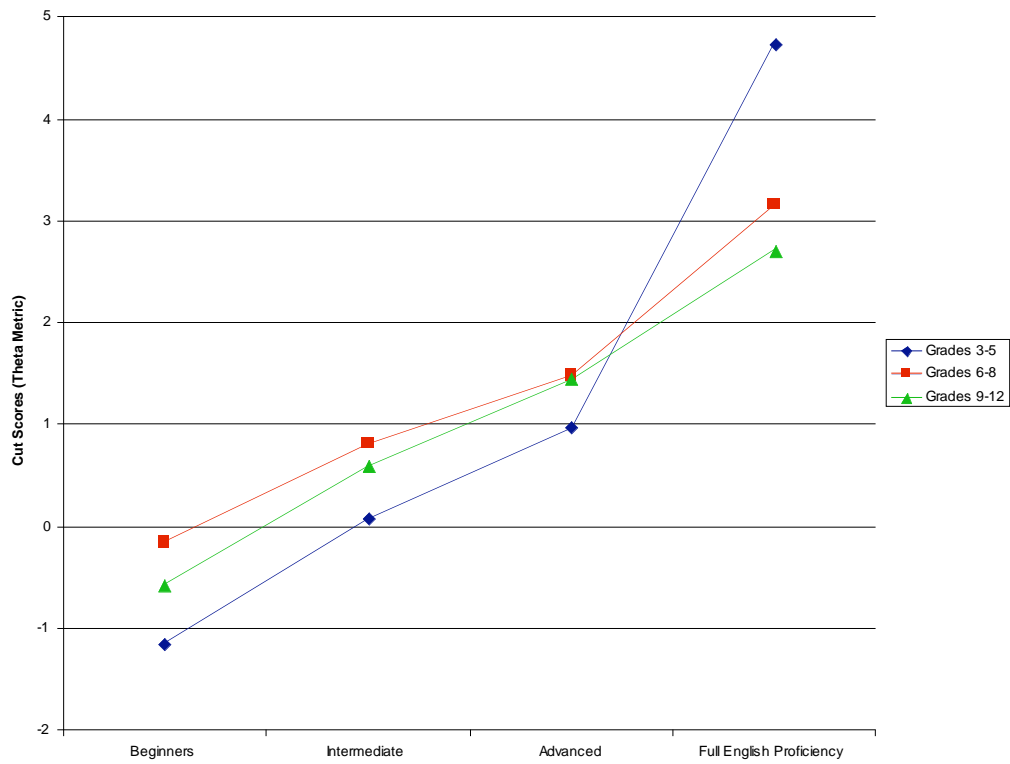


Figure 4. ELDA Cut Scores in Theta Metric – Writing



Need for a Short Version of the ELDA

As noted previously, to measure a wide range of English language proficiencies, the ELDA operational forms include many items and require substantial test administration time. The number of items in the ELDA operational test forms range from 50 to 60 for the Reading and Listening tests, depending on the grade cluster. Across the three grade clusters, the Writing tests comprise from 16 to 23 items in the first operational form. The first operational Speaking test forms contained 16 items across all three grade clusters. Administration of the ELDA test battery is officially untimed; however, general guidelines for times are provided in the Administration Manual as indicated in Table 3.

Table 3. General Test Administration in Time Guidelines

Grade Cluster	Listening	Speaking	Reading	Writing
3–5	1 hour, 20 mins	25 mins	1 hour	1 hour
6–8	1 hour, 20 mins	25 mins	1 hour	1 hour
9–12	1 hour, 40 mins	25 mins	1 hour, 15 mins	1 hour

As Table 3 illustrates, test administration times for the ELDA are rather long, requiring approximately 4 hours for students to complete the entire battery. The Ohio Department of Education (ODE) sought to develop a shorter version of the ELDA to ease the administration burdens on schools and students. AIR proposed to dramatically reduce the length of the ELDA tests while still providing accurate and reliable information about student English language proficiency in Listening, Speaking, Reading, and Writing.

Performance results from the first operational administration of the ELDA indicated that the majority of the student population to whom the test was administered scored in the upper ranges of the raw score distribution. The figures in Appendix B present the raw score frequency distributions for each subject and grade cluster assessed in the ELDA.

As Figure B1 illustrates, for example, student performance on the grade cluster 3–5 Listening test was highly negatively skewed, with 80% of students scoring at least 34 points and 50% of students scoring at least 43 points out of 50 available points. More than 60% of the test items are used to assess less than 20% of the LEP student population. Similar results were obtained for Listening in the 6–8 and 9–12 grade clusters as seen in Figure B2 and Figure B3, respectively.

Student performance on the Speaking test was also highly negatively skewed, as shown in Figures B4–B6. In the case of Speaking, 80% of students in the 3–5 and 6–8 grade clusters achieved 24 out of 32 points, and 50% of students at all three grade clusters received 29 or more points out of 32 possible points. The ELDA operational speaking test therefore uses about 75% of available score points to characterize less than 20% of the student population, and it uses only about 4 raw score points to discriminate among students in the top half of the distribution.

Although not nearly as dramatic, student performance in Reading was also negatively skewed, with much of the test information targeted toward the most limited English language learners. About 80% of students achieved 40% of available raw score points on the Reading test. The situation in Writing is similar, with about 80% of students achieving 40–50% of available score points, depending on the grade cluster.

These results indicated that the ELDA operational forms could be shortened dramatically by eliminating the easiest items in the operational item bank while maintaining a proportional representation of items across content standards within each subject area. In some cases, the easiest items within a subject area are concentrated within specific content standards or topic areas. When this proved to be the case, ODE opted to maintain breadth of content coverage rather than simply maximize form difficulty. By reducing the number of easy items, the overall difficulty of the tests increased so that the tests are less heavily targeted toward the beginning-level skills. Shortening the length of the ELDA tests also, however, results in decreased reliability of the obtained test score. Thus, although our goal was to shorten the ELDA tests as much as possible, we also had to be mindful of estimated form reliabilities to determine the appropriate number of items to include in each test form.

In sum, the goals of OTELA are (1) to reduce test administration time by reducing the number of items on the test, (2) to reduce emphasis on beginning/entry level skills and (3) to produce reliable and valid test scores. In the next section, we present the test development framework for the OTELA.

2. OTELA Test Development Framework

Test Specifications. ODE contracted with AIR to develop two reduced-length ELDA forms per subject and grade cluster. Items were to be selected on the basis of their psychometric properties, their contribution to measurement at key points on the scale (such as the proficient cut score) and their content coverage. These activities resulted in the development of 24 reduced-length ELDA forms (three grade clusters \times four subjects \times two forms), 12 of which will be administered in Ohio during April 2006. The number of items and administration time estimates (about 1.5 hours per student across all four subjects) are shown in Table 4. The OTELA Listening and Reading forms are made up of approximately one-third of the number of items making up a standard ELDA operational form, and the OTELA Speaking forms contain three-fourths the number of items. It was not possible to maintain adequate levels of form reliability for a shorter form. The OTELA Writing forms are approximately half the length of standard ELDA operational forms.

Table 4. OTELA Test Administration Times and Number of Items per Form

Time in minutes	Subjects	Number of Items		
		Grades 3–5	Grades 6–8	Grades 9–12
20	Listening	18	18	20
20	Speaking	12	12	12
20	Reading	20	20	20
30	Writing	9 MC 2 CR	9 MC 2 CR	9 MC 3 CR

To identify a target average item difficulty estimate for each OTELA test form, we computed an estimated difficulty on the basis of the most difficult items among the current ELDA operational tests and the estimated percentage of students whose ability is at or below the target difficulty. For the Reading and Speaking tests, we first computed the median item difficulty estimate for the top half most difficult items in each of the three ELDA operational test forms currently defined at each grade cluster. The average of the median item difficulty estimates across the three operational forms at each grade cluster constituted the target difficulty for the corresponding OTELA test forms. For the Writing tests, the two most difficult constructed-response items and the six most difficult multiple-choice items were used to estimate the target difficulty so that both multiple-choice and constructed-response items were taken into account. For the Listening tests, the most difficult 18 (3–5 grade cluster) or 20 (6–8 and 9–12 grade clusters) items were used because a target based on the top half most difficult items resulted in an overly easy set of test forms. The resulting target average item difficulty estimates for each OTELA test form are presented in Table 5.

Table 5. Target Average Item Difficulty Estimates for the OTLEA Tests

Grade Cluster	Listening	Reading	Speaking	Writing
3–5	0.41	0.33	0.35	0.33
6–8	0.83	0.54	0.58	0.63
9–12	1.19	0.94	1.10	0.73

Reduction of ELDA Item Pool. Because a major goal of the OTLEA tests is to reduce test length by eliminating the easiest items in the operational test forms, we began the test construction process by eliminating the easiest items in the ELDA item pool. On the basis of the frequency distributions of item difficulty parameters in Reading and Listening, we eliminated from consideration items with Rasch difficulty parameter estimates less than -1.8 in the 3–5 grade cluster, less than -1.4 in the 6–8 grade cluster and less than -0.5 in the 9–12 grade cluster. For Speaking and Writing, however, no items were excluded from the item pools. In Speaking, items are constructed in sets that demand increasing levels of speaking proficiency (e.g., connect, tell, expand, reason). The pool of Writing items was not reduced because there are relatively few items in the pool, and most of them are of medium or high difficulty.

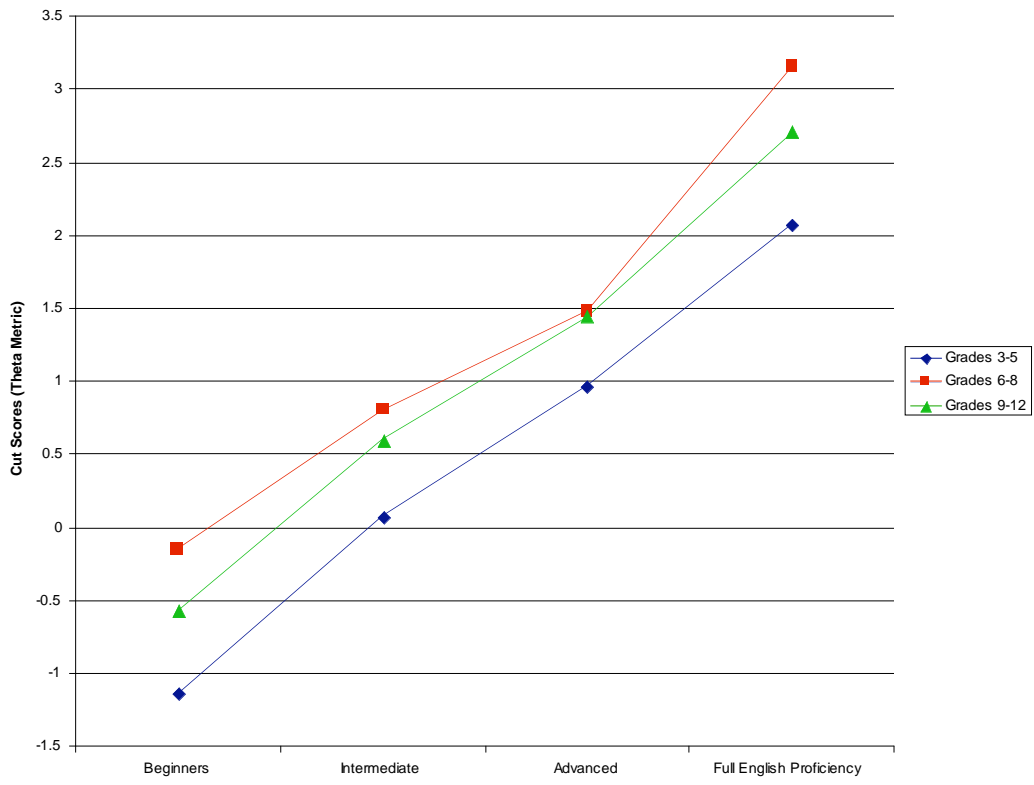
Form Construction. To assemble the OTELA forms, AIR content specialists worked with AIR psychometricians to ensure that each form balances the content and statistical requirements of the OTELA. Twenty-four OTELA forms were constructed. The next section details the test characteristics of the new OTELA forms.

3. OTELA Test Characteristics

OTELA Performance Standards. The OTELA is designed provide student assessment results that are fully comparable with assessment results from the ELDA. To achieve this goal, the OTELA utilizes the same performance standards adopted by the LEP-SCASS for the ELDA, as established through a standard setting process conducted by Measurement Inc. A consequence of this decision is that cut scores are not uniformly well articulated across grade level clusters. As noted previously, the performance standards for listening are similar for students in both the 6-8 and 9-12 grade clusters, and for writing, performance standards for students in the grade 6-8 grade cluster are uniformly higher than for students in the grade 9-12 grade cluster.

In the process of adopting ELDA performance standards for the OTELA, ODE, in consultation with the Ohio LEP Advisory Committee, elected to revise one ELDA performance level cut score. In the ELDA performance standards for writing, students in the grade 3-5 cluster must substantially outperform students in both the 6-8 and 9-12 grade clusters to achieve Full English Proficiency. This particular cut score results in almost no grade 3-5 students achieving Full English Proficiency in writing and seems generally inconsistent with the placement of other cut scores for writing in the 3-5 grade cluster. To address this issue, we used a linear regression approach to identify a cut score for Full English Proficiency at the grade 3-5 cluster from the cut scores identified for Beginning, Intermediate, and Advanced performance levels on the grade 3-5 writing assessment. This analysis identified a cut score of 2.08 (in the theta metric) for the Full English Proficiency cut score at the 3-5 grade cluster. This cut score and estimated impact data for the revised performance standard was submitted to the Ohio LEP Advisory Committee for their consideration. The Ohio LEP Advisory Committee recommended that ODE adopt the revised performance standard, which ODE has done. The placement of the revised cut score is illustrated in Figure 4 and is used to define Full English Proficiency in writing for the 3-5 grade cluster conversion tables.

Figure 4. OTELA Cut Scores in Theta Metric – Revised Writing



Conversion Tables. As noted previously, OTELA test scores are reported on the ELDA scales. (Appendix A contains the theta-to-ELDA scaled score transformations for each subject.) Table 6 shows the OTELA cut scores for each proficiency level on the ELDA scale. Appendix C provides the final transformations of raw scores to Rasch ability estimates (in logit measure) to ELDA scaled scores. The tables in Appendix C also list the error of estimation for each value, as well as the proficiency level associated with each score point. Theta estimates for extreme scores (i.e., 0 and perfect) were estimated using a linear extrapolation procedure in which $\theta(0) = 2 \times \theta(1) - \theta(2)$ and $\theta(n) = 2 \times \theta(n - 1) - \theta(n - 2)$. This procedure produces values very close to the values obtained by adding or subtracting .5 from zero or perfect scores, respectively, as recommended by Berkson (as cited in Linacre, 2004).

Table 6. OTELA Cut Scores on the ELDA Scale

Subject	Grade Level	Pre-functional	Beginners	Intermediate	Advanced	Full English Proficiency
Reading	3–5	Below 450	450 – 579	580 – 647	648 – 769	770 and Above
Reading	6–8	Below 460	460 – 611	612 – 689	690 – 828	829 and Above
Reading	9–12	Below 545	545 – 629	630 – 717	718 – 849	850 and Above
Listening	3–5	Below 450	450 – 543	544 – 644	645 – 724	725 and Above
Listening	6–8	Below 554	554 – 625	626 – 717	718 – 805	806 and Above
Listening	9–12	Below 556	556 – 631	632 – 728	729 – 849	850 and Above
Speaking	3–5	Below 450	450 – 546	547 – 667	668 – 808	809 and Above
Speaking	6–8	Below 458	458 – 610	611 – 718	719 – 824	825 and Above
Speaking	9–12	Below 570	570 – 649	650 – 764	765 – 849	850 and Above
Writing	3–5	Below 450	450 – 576	577 – 668	669 – 866	867 and Above
Writing	6–8	Below 553	553 – 652	653 – 721	722 – 893	894 and Above
Writing	9–12	Below 509	509 – 630	631 – 718	719 – 849	850 and Above

Form Difficulties. A summary of the average item difficulty and total score points for each OTELA form is provided in Table 6. Examination of Table 7 reveals that the realized average item difficulty estimates for the OTELA forms are very close to the target difficulties for the Reading, Listening and Writing tests identified in Table 5. For Speaking, two sets of grade 3–5 Speaking forms were proposed to ODE. The first set of forms sought to maximize average form difficulty (with average item difficulties near the target), resulting in a lack of topic area coverage. The second set of forms explicitly maximized topic area coverage at the expense of average form difficulty. Consistent with its goal of maximizing content/topic area coverage, ODE approved the set of Speaking forms that maximized topic area coverage.

Table 7. Average Item Difficulty Estimates for the OTELA Tests

Subject	Form	Grades 3–5		Grades 6–8		Grades 9–12	
		Score Points	Average Difficulty	Score Points	Average Difficulty	Score Points	Average Difficulty
Reading	ELDA OP Form 1	50	–0.38	50	–0.1	60	0.41
	OTELA Form A	20	0.35	20	0.53	20	0.87
	OTELA Form B	20	0.35	20	0.53	20	0.88
Listening	ELDA OP Form 1	50	–0.54	50	0.03	60	0.3
	OTELA Form A	18	0.42	18	0.8	20	1.1
	OTELA Form B	18	0.41	18	0.8	20	1.08
Speaking	ELDA OP Form 1	32	–0.36	32	–0.17	32	0.36
	OTELA - Form A	24	0.14	24	0.60	24	0.91
	OTELA - Form B	24	0.11	24	0.58	24	0.91
Writing	ELDA OP Form 1	25	–0.21	31	0.11	34	0.04
	OTELA Form A	17	0.25	17	0.43	20	0.80
	OTELA Form B	17	0.25	17	0.45	20	0.78

Estimated Form Reliabilities. Table 8 displays the estimated form reliabilities based on the Spearman Brown prophecy formula for the OTELA tests as well as the coefficient alpha reliability estimates from the first operational administration of the ELDA (operational Form 1). The reliability estimates for the OTELA forms are moderately high, ranging from 0.76 to 0.91, although the reduced-length tests resulted in slight decreases in reliability estimates compared with those of the original ELDA tests.

Table 8. Estimated Form Reliabilities for the OTELA Tests

		Grade 3–5	Grade 6–8	Grade 9–12
Subject	Form	Estimated Reliability	Estimated Reliability	Estimated Reliability
Reading	ELDA OP Form 1	0.94	0.92	0.94
	OTELA Form A	0.86	0.82	0.84
	OTELA Form B	0.86	0.82	0.84
Listening	ELDA OP Form 1	0.94	0.95	0.95
	OTELA Form A	0.85	0.87	0.86
	OTELA Form B	0.85	0.87	0.86
Speaking	ELDA OP Form 1	0.91	0.94	0.93
	OTELA - Form A	0.88	0.92	0.91
	OTELA - Form B	0.88	0.92	0.91
Writing	ELDA OP Form 1	0.82	0.87	0.89
	OTELA Form A	0.76	0.79	0.83
	OTELA Form B	0.76	0.79	0.83

Estimated Performance Level Outcomes. After obtaining the relationship between raw and scaled scores, we estimated the percentage of students at each test performance level. Table 9 presents the estimated percentage of students “at” each test performance level for OTELA Form A on the basis of projections from the results of the first operational administration of ELDA (operational Form 1).

Table 9. Estimated Percentage of Students at Each Performance Level

Estimated Percentage of Students at Each Performance Level – OTELA Form A						
Subject	Grade Cluster	Performance Level	Raw Score Performance Standard	Scaled Score Performance Standard	Percentage	
Reading	3–5	Pre-functional			16.24	
		Beginners	6.00	454	26.13	
		Intermediate	11.00	595	10.81	
		Advanced	13.00	651	25.93	
		Full English Proficiency	17.00	797	20.88	
	6–8	Pre-functional				13.38
		Beginners	6.00	478	29.11	
		Intermediate	11.00	618	19.72	
		Advanced	14.00	704	27.04	
		Full English Proficiency	18.00	877	10.75	
	9–12	Pre-functional				20.77
		Beginners	7.00	551	17.01	
		Intermediate	10.00	633	22.18	
		Advanced	14.00	745	18.73	
		Full English Proficiency	17.00	859	21.31	
Listening	3–5	Pre-functional			7.34	
		Beginners	5.00	473	8.06	
		Intermediate	8.00	570	13.01	
		Advanced	11.00	660	13.93	
		Full English Proficiency	13.00	726	57.66	
	6–8	Pre-functional				12.66
		Beginners	7.00	585	5.57	
		Intermediate	9.00	644	12.12	
		Advanced	12.00	736	12.69	
		Full English Proficiency	14.00	808	56.96	
	9–12	Pre-functional				10.17
		Beginners	6.00	567	8.76	
		Intermediate	9.00	654	11.77	
		Advanced	12.00	734	22.22	
		Full English Proficiency	16.00	862	47.08	
Speaking	3–5	Pre-functional			4.40	
		Beginners	6.00	473	3.28	
		Intermediate	10.00	552	13.43	
		Advanced	17.00	685	32.65	
		Full English Proficiency	22.00	840	46.24	
	6–8	Pre-functional				4.66
		Beginners	3.00	459	8.62	
		Intermediate	12.00	624	10.44	

Estimated Percentage of Students at Each Performance Level – OTELA Form A					
Subject	Grade Cluster	Performance Level	Raw Score Performance Standard	Scaled Score Performance Standard	Percentage
		Advanced	18.00	729	23.30
		Full English Proficiency	22.00	866	52.98
	9–12	Pre-functional			10.59
		Beginners	8.00	589	6.46
		Intermediate	12.00	662	16.40
		Advanced	18.00	775	16.31
		Full English Proficiency	21.00	857	50.24
Writing	3–5	Pre-functional			12.42
		Beginners	5.00	473	33.33
		Intermediate	9.00	594	31.64
		Advanced	12.00	702	14.89
		Full English Proficiency	14.00	811	7.72
	6–8	Pre-functional			17.33
		Beginners	7.00	571	20.08
		Intermediate	10.00	668	19.32
		Advanced	12.00	745	34.24
		Full English Proficiency	15.00	903	9.04
	9–12	Pre-functional			12.69
		Beginners	6.00	533	18.82
		Intermediate	10.00	635	22.66
		Advanced	13.00	724	30.20
		Full English Proficiency	16.00	857	15.63

Classification Consistency Estimates. Table 10 presents estimated OTELA Form A classification consistency estimates at each of the performance standard cut scores as projected from 2005 ELDA operational test administration data. Classification consistency indexes the agreement between the classification resulting from students' observed scores and the classification resulting from scores as projected from a hypothetical independent administration of a parallel test form. Classification consistency estimates are derived following Huynh's (1979) use of the beta-binomial model. Kappa provides an index of classification consistency that is corrected for chance levels of agreement.

Table 10. Estimated Classification Consistency Estimates for the OTELA Tests

Estimated Classification Consistency Estimates – OTELA Form A				
Subject	Grade Cluster	Proficiency	Consistency	$\kappa_{\text{Consistency}}$
Reading	3–5	Beginners	0.90	0.61
		Intermediate	0.84	0.67
		Advanced	0.84	0.67
		Full English Proficiency	0.88	0.61
	6–8	Beginners	0.90	0.55
		Intermediate	0.82	0.63
		Advanced	0.82	0.62
		Full English Proficiency	0.91	0.50
	9–12	Beginners	0.89	0.62
		Intermediate	0.85	0.66
		Advanced	0.83	0.65
		Full English Proficiency	0.87	0.60
Listening	3–5	Beginners	0.95	0.53
		Intermediate	0.90	0.61
		Advanced	0.85	0.65
		Full English Proficiency	0.83	0.66
	6–8	Beginners	0.93	0.64
		Intermediate	0.90	0.67
		Advanced	0.87	0.70
		Full English Proficiency	0.85	0.70
	9–12	Beginners	0.94	0.62
		Intermediate	0.91	0.68
		Advanced	0.87	0.71
		Full English Proficiency	0.85	0.70
Speaking	3–5	Beginners	0.98	0.62
		Intermediate	0.96	0.70
		Advanced	0.91	0.76
		Full English Proficiency	0.87	0.74
	6–8	Beginners	0.98	0.65
		Intermediate	0.95	0.80
		Advanced	0.93	0.82
		Full English Proficiency	0.90	0.80
	9–12	Beginners	0.95	0.75
		Intermediate	0.94	0.79
		Advanced	0.91	0.80
		Full English Proficiency	0.89	0.78
Writing	3–5	Beginners	0.87	0.40
		Intermediate	0.75	0.50
		Advanced	0.81	0.46
		Full English Proficiency	0.90	0.37
	6–8	Beginners	0.86	0.48
		Intermediate	0.77	0.53
		Advanced	0.77	0.52
		Full English Proficiency	0.88	0.42

Estimated Classification Consistency Estimates – OTELA Form A				
Subject	Grade Cluster	Proficiency	Consistency	κConsistency
	9–12	Beginners	0.90	0.47
		Intermediate	0.80	0.56
		Advanced	0.79	0.56
		Full English Proficiency	0.86	0.50

References

- Huynh, H. (1979). Statistical inference for two reliability indices in mastery testing based on the beta-binomial model. *Journal of Educational Statistics*, 4, 231–246.
- Kolen, M. J. & Brennan, R. L. (1995). *Test equating: Methods and Practices*. New York: Springer-Verlag.
- Linacre, J. M. (2004). *A user's guide to Winsteps*. Chicago: MESA Press.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

APPENDIX A

Theta to ELDA Scaled Score Transformations

Table A1. Theta to ELDA Scaled Score Transformations

Subject		Test Location	Linear Transformation
Reading			$SS = 450 + 121.9512 \times (\text{Theta} + .62)$
	Lower Trim	If $\text{Theta} < -2.987$	$SS = 100 + 23.81152 \times (\text{Theta} + 5.627)$
	Upper Trim	If $\text{Theta} > 3.021$	$SS = 894 + 22.96023 \times (\text{Theta} - 3.021)$
Listening			$SS = 450 + 119.0476 \times (\text{Theta} + .83)$
	Lower Trim	If $\text{Theta} < -3.21$	$SS = 100 + 30.29 \times (\text{Theta} + 5.42)$
	Upper Trim	If $\text{Theta} > 3.00$	$SS = 906 + 17.81 \times (\text{Theta} - 3.00)$
Speaking			$SS = 450 + 91.95402 \times (\text{Theta} + 1.39)$
	No Lower Trim		
	Upper Trim	If $\text{Theta} > 3.5$	$SS = 900 + 32.051 \times (\text{Theta} - 3.5)$
Writing			$SS = 450 + 103.8961 \times (\text{Theta} + 1.14)$
	No Lower Trim		
	Upper Trim	If $\text{Theta} > 3.12$	$SS = 893 + 25.333 \times (\text{Theta} - 3.12)$

APPENDIX B

Raw Score Frequency Distributions for the 2005 ELDA Operational Test Forms

Figure B1. Distribution of Student Raw Score Performance — Grades 3–5 Listening

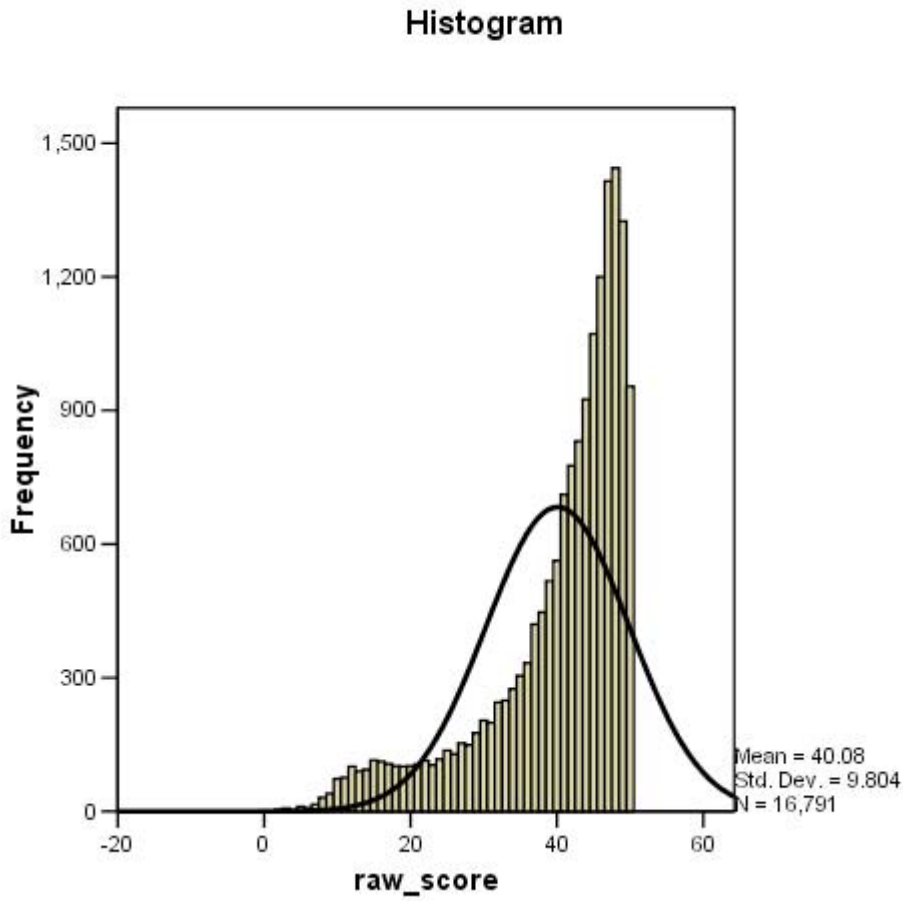


Figure B2. Distribution of Student Raw Score Performance — Grades 6–8 Listening

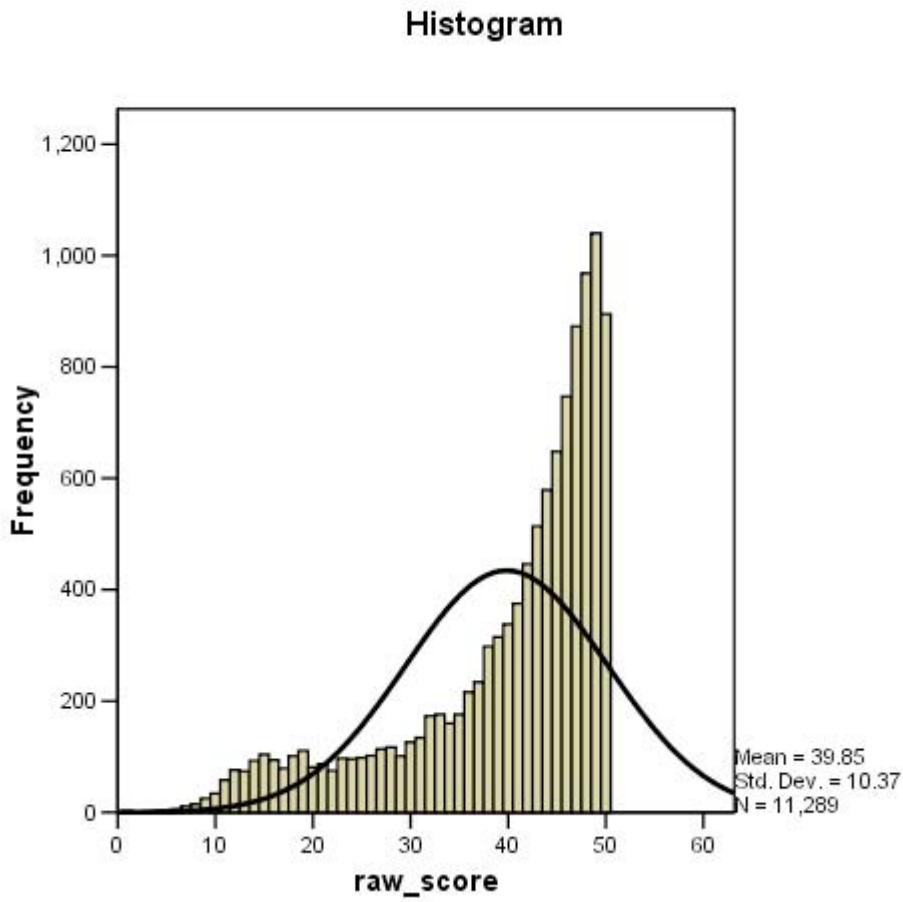


Figure B3. Distribution of Student Raw Score Performance — Grades 9–12 Listening

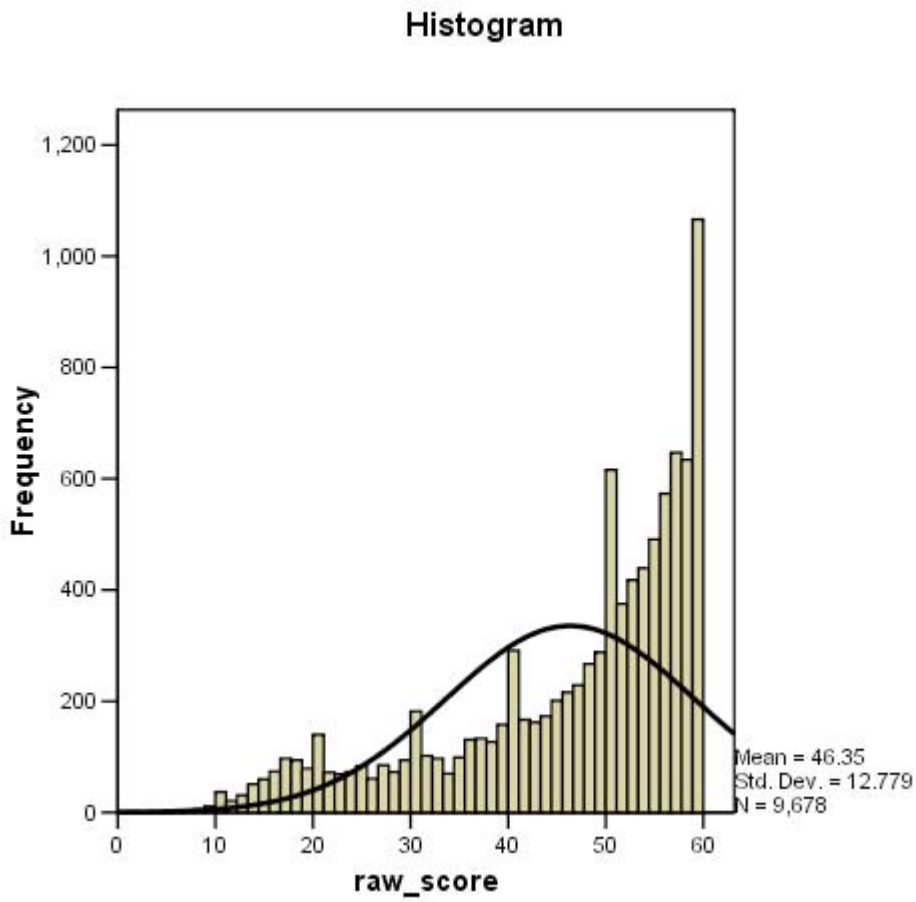


Figure 4. Distribution of Student Raw Score Performance — Grades 3–5 Speaking

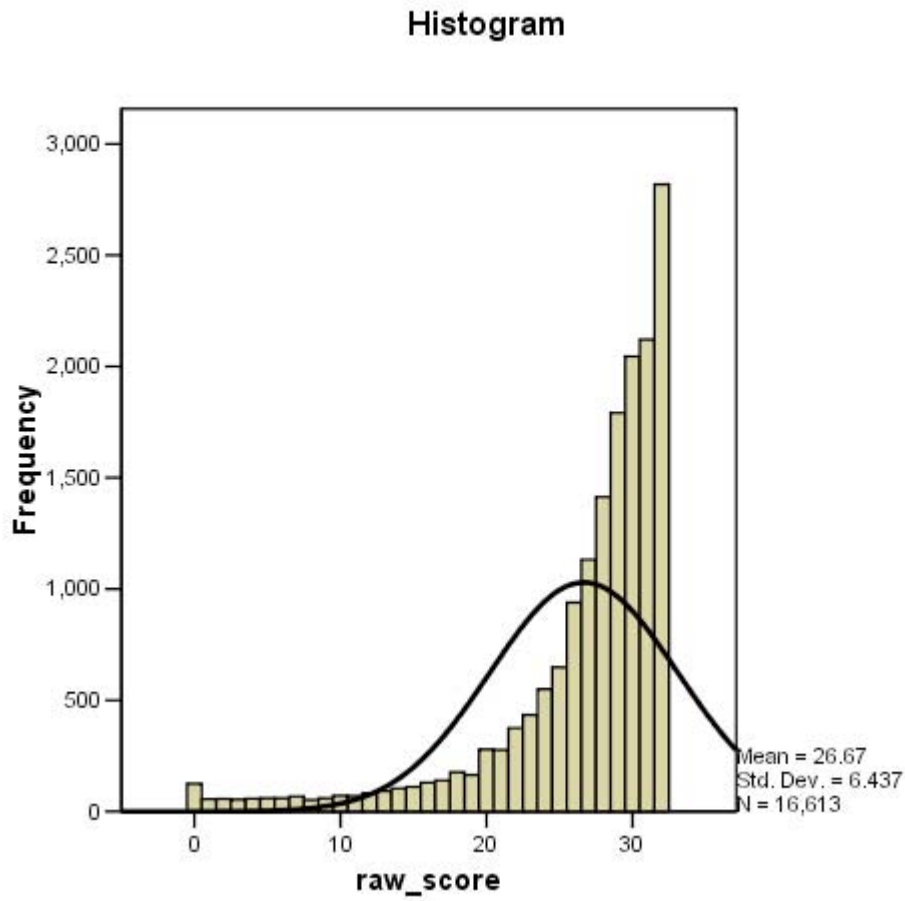


Figure 5. Distribution of Student Raw Score Performance — Grades 6–8 Speaking

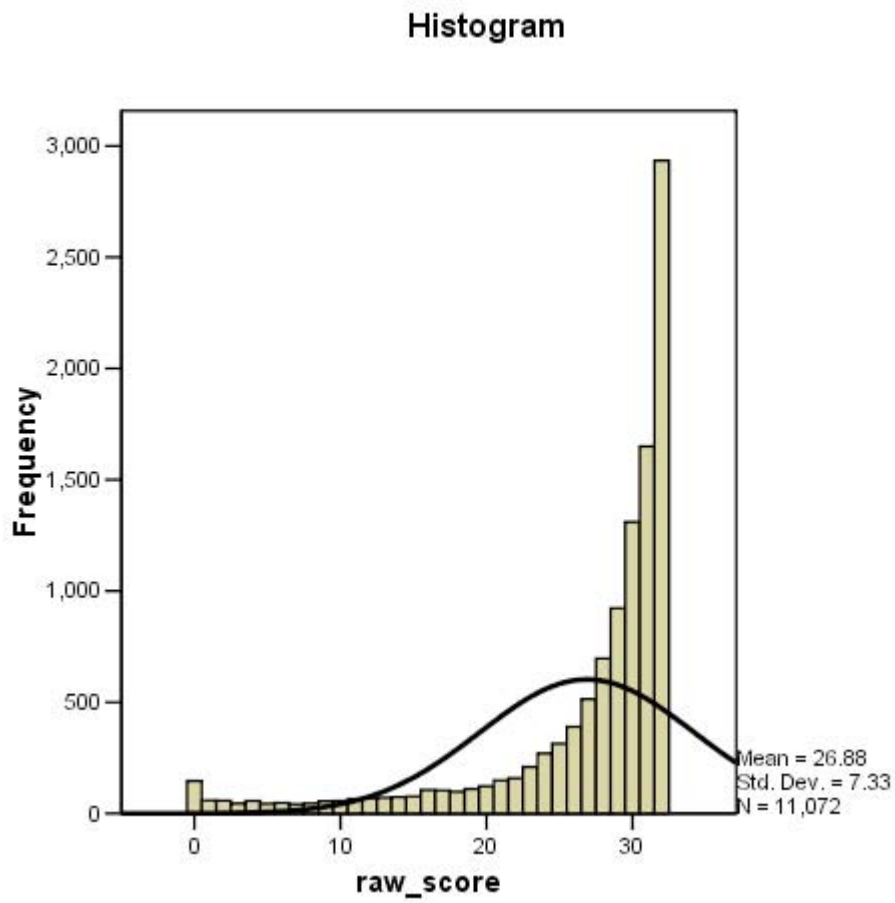


Figure 6. Distribution of Student Raw Score Performance — Grades 9–12 Speaking

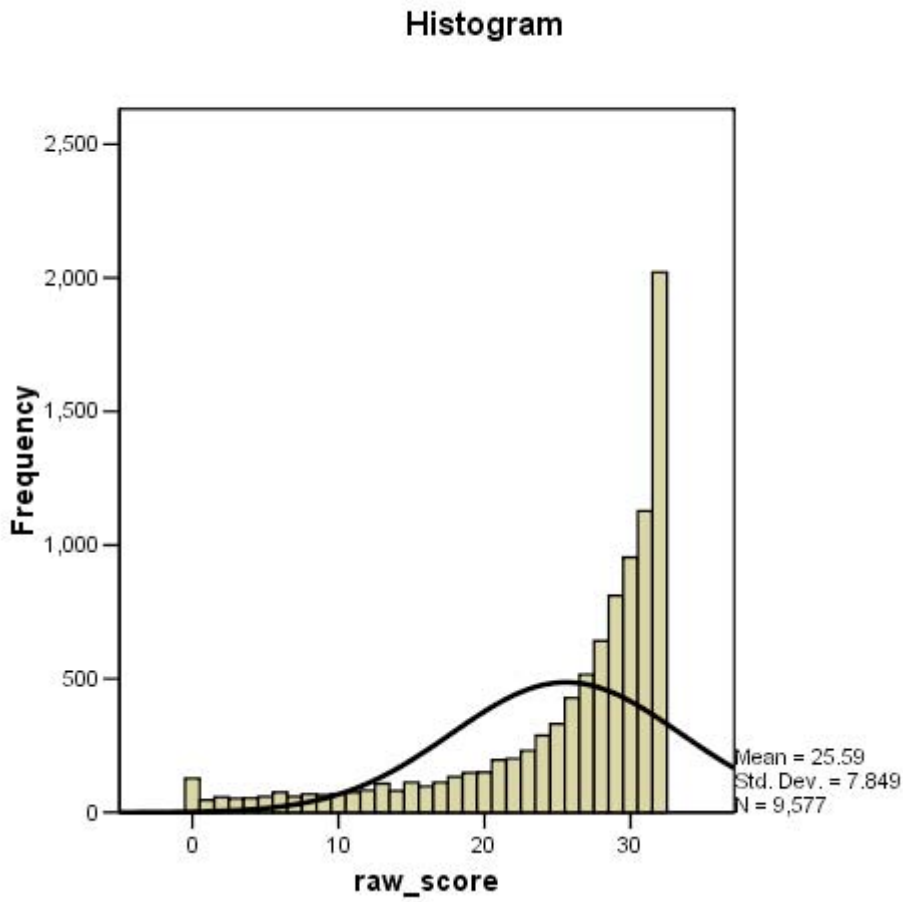


Figure 7. Distribution of Student Raw Score Performance — Grades 3–5 Reading

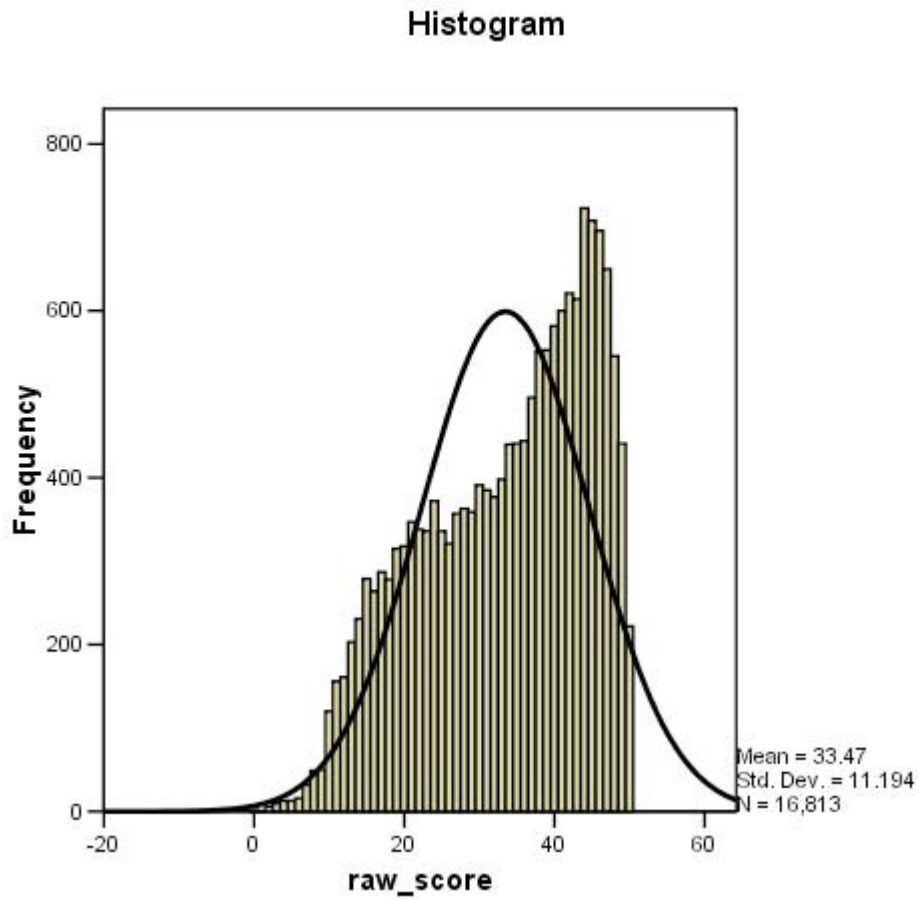


Figure 8. Distribution of Student Raw Score Performance — Grades 6–8 Reading

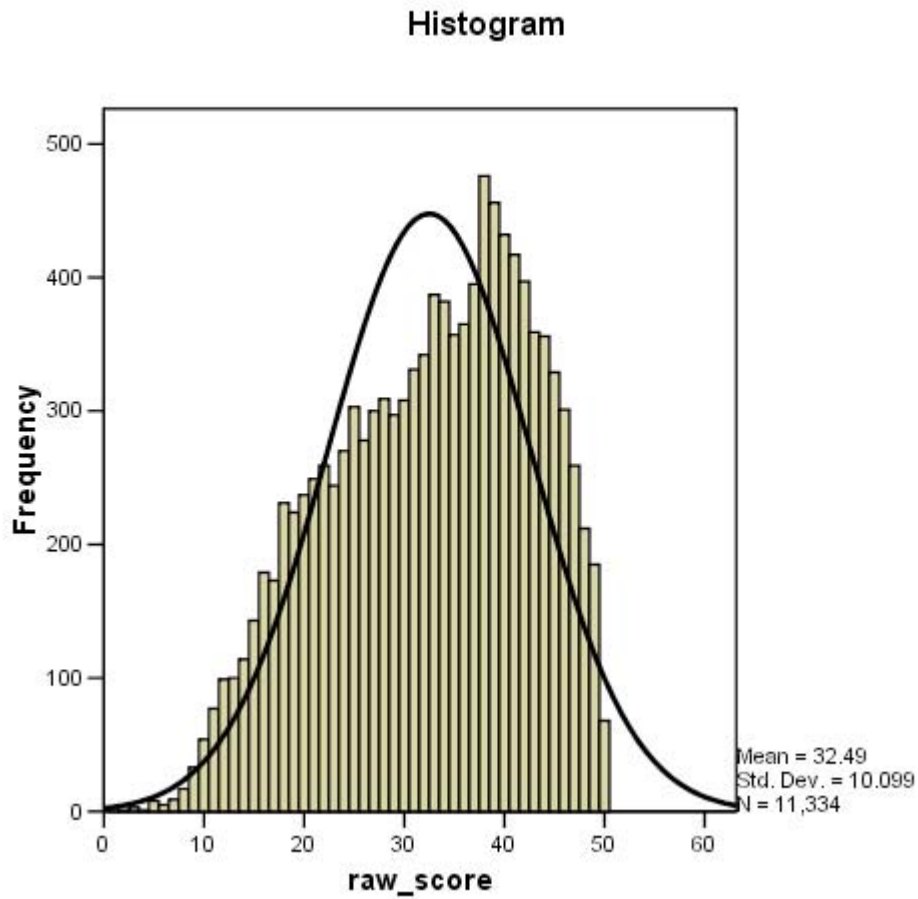


Figure 9. Distribution of Student Raw Score Performance — Grades 9–12 Reading

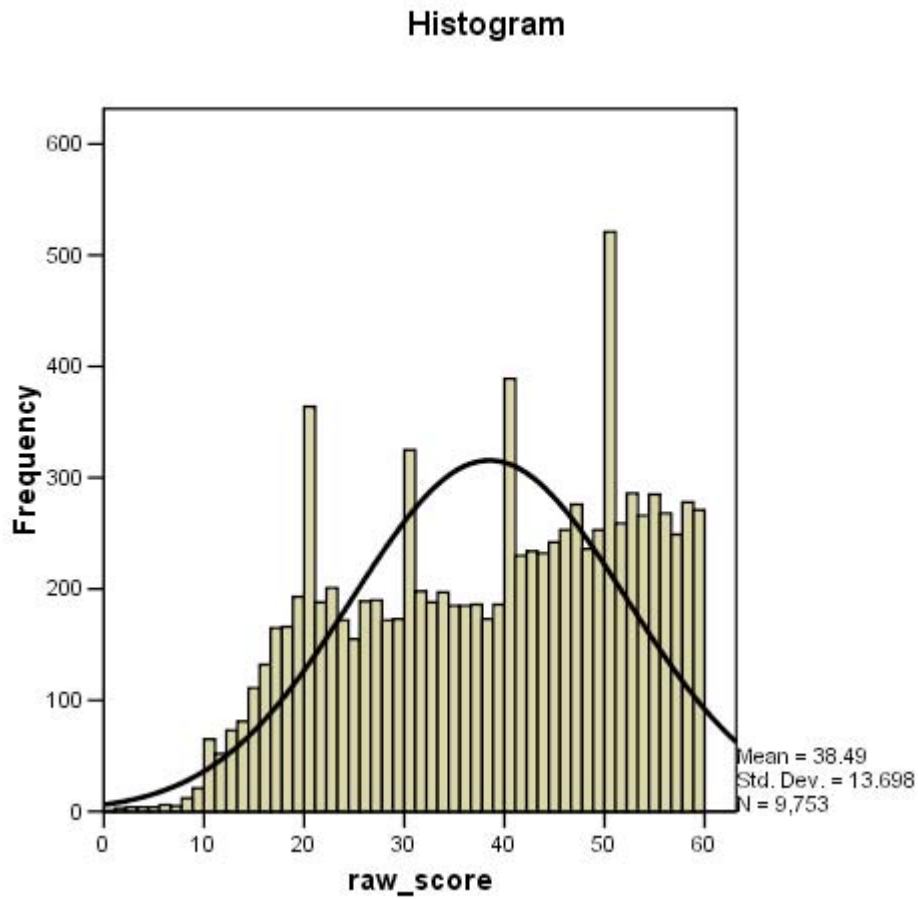


Figure 10. Distribution of Student Raw Score Performance — Grades 3–5 Writing

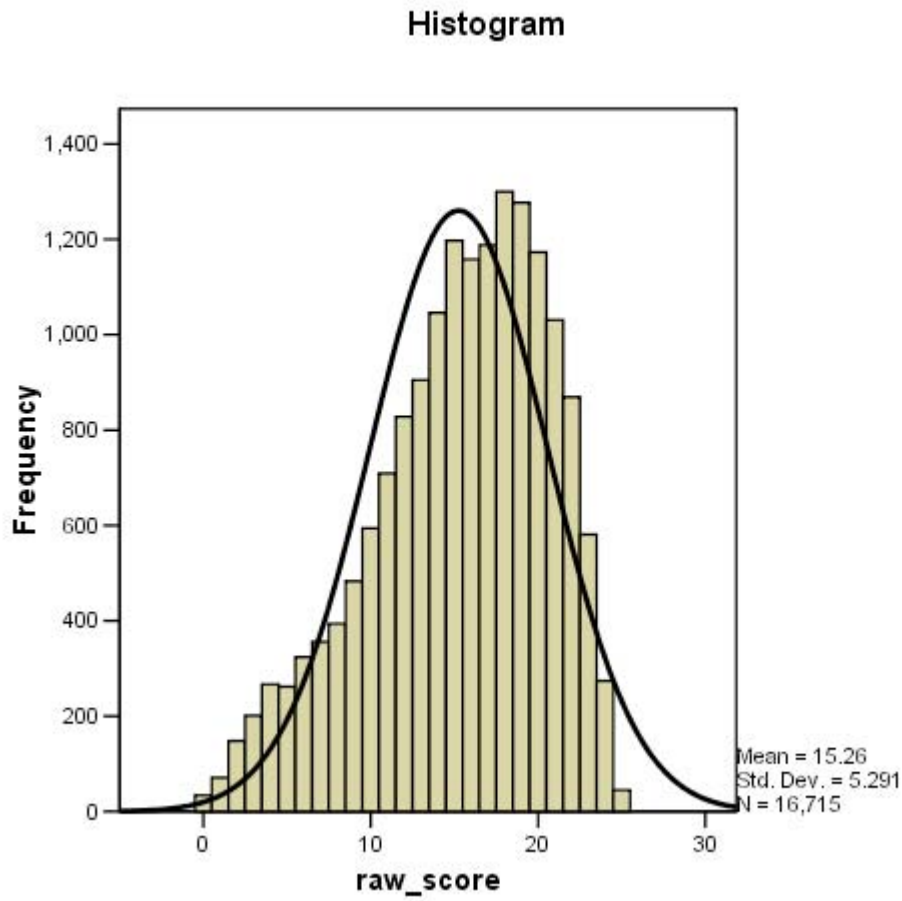


Figure 11. Distribution of Student Raw Score Performance — Grades 6–8 Writing

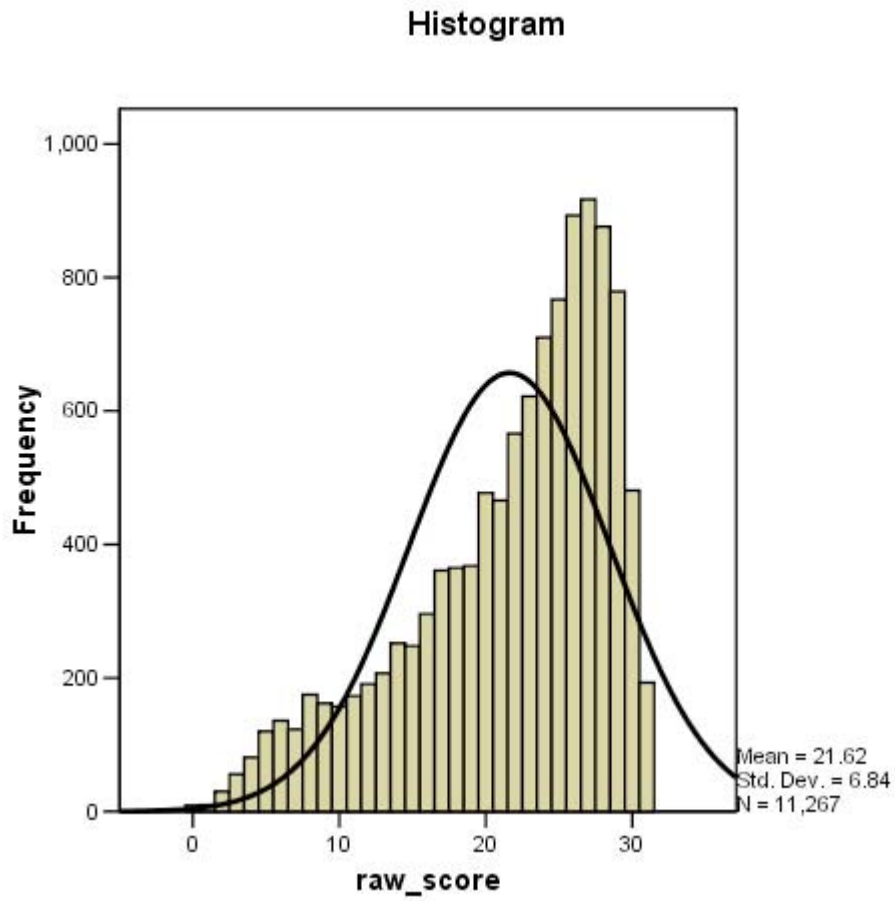
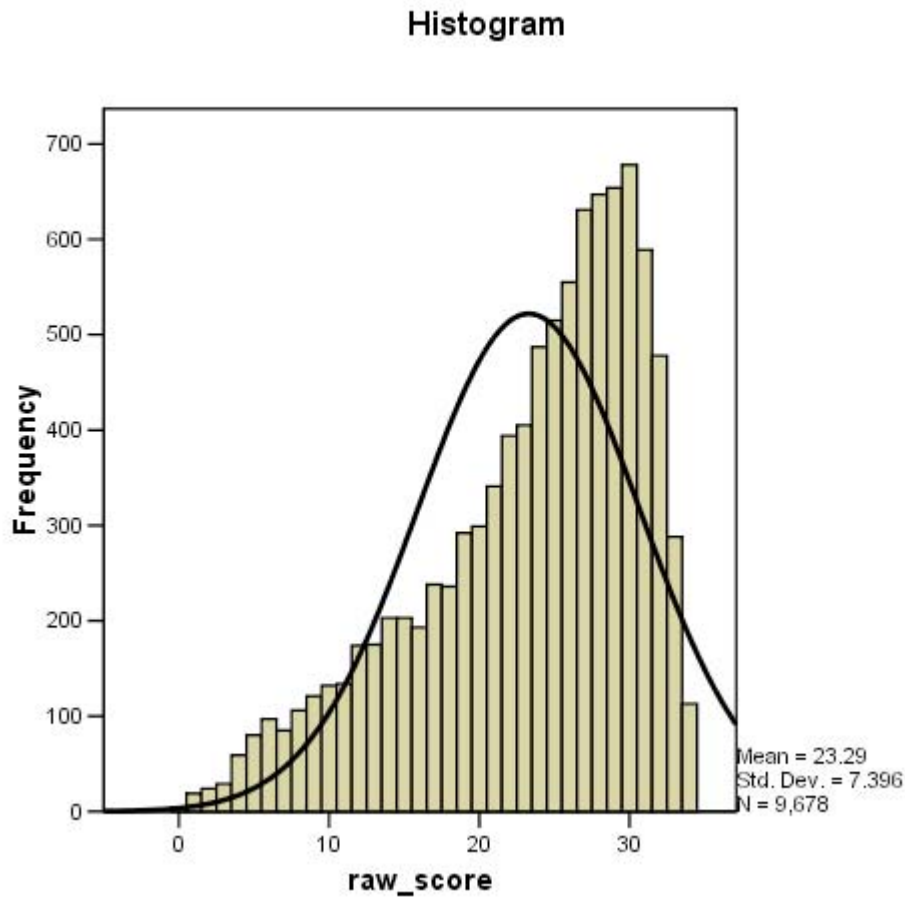


Figure 12. Distribution of Student Raw Score Performance — Grades 9–12 Writing



APPENDIX C

OTELA Conversion Tables

Table C1. Conversion Tables — Grades 3–5 Reading Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.57	1.47	149	Level 1	Pre-functional
1.00	-2.80	1.04	184	Level 1	Pre-functional
2.00	-2.03	0.76	278	Level 1	Pre-functional
3.00	-1.54	0.64	338	Level 1	Pre-functional
4.00	-1.17	0.58	383	Level 1	Pre-functional
5.00	-0.86	0.54	421	Level 1	Pre-functional
6.00	-0.58	0.51	454	Level 2	Beginners
7.00	-0.33	0.49	485	Level 2	Beginners
8.00	-0.10	0.48	514	Level 2	Beginners
9.00	0.13	0.47	541	Level 2	Beginners
10.00	0.35	0.47	568	Level 2	Beginners
11.00	0.57	0.47	595	Level 3	Intermediate
12.00	0.80	0.48	623	Level 3	Intermediate
13.00	1.03	0.49	651	Level 4	Advanced
14.00	1.28	0.51	682	Level 4	Advanced
15.00	1.55	0.54	715	Level 4	Advanced
16.00	1.86	0.58	752	Level 4	Advanced
17.00	2.23	0.64	797	Level 5	Full English Proficiency
18.00	2.71	0.76	856	Level 5	Full English Proficiency
19.00	3.48	1.04	905	Level 5	Full English Proficiency
20.00	4.25	1.47	922	Level 5	Full English Proficiency

Table C2. Conversion Table — Grades 3–5 Reading Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.53	1.46	150	Level 1	Pre-functional
1.00	-2.76	1.04	189	Level 1	Pre-functional
2.00	-2.00	0.76	282	Level 1	Pre-functional
3.00	-1.51	0.64	341	Level 1	Pre-functional
4.00	-1.15	0.57	386	Level 1	Pre-functional
5.00	-0.84	0.53	423	Level 1	Pre-functional
6.00	-0.57	0.51	456	Level 2	Beginners
7.00	-0.33	0.49	486	Level 2	Beginners
8.00	-0.10	0.48	514	Level 2	Beginners
9.00	0.13	0.47	541	Level 2	Beginners
10.00	0.35	0.47	568	Level 2	Beginners
11.00	0.57	0.47	595	Level 3	Intermediate
12.00	0.79	0.48	622	Level 3	Intermediate
13.00	1.02	0.49	650	Level 4	Advanced
14.00	1.27	0.51	680	Level 4	Advanced
15.00	1.54	0.54	713	Level 4	Advanced
16.00	1.85	0.58	751	Level 4	Advanced
17.00	2.22	0.64	796	Level 5	Full English Proficiency
18.00	2.70	0.76	855	Level 5	Full English Proficiency
19.00	3.47	1.04	904	Level 5	Full English Proficiency
20.00	4.24	1.47	922	Level 5	Full English Proficiency

Table C3. Conversion Table — Grades 6–8 Reading Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.39	1.47	153	Level 1	Pre-functional
1.00	-2.61	1.04	207	Level 1	Pre-functional
2.00	-1.83	0.76	302	Level 1	Pre-functional
3.00	-1.34	0.65	362	Level 1	Pre-functional
4.00	-0.97	0.58	407	Level 1	Pre-functional
5.00	-0.66	0.54	445	Level 1	Pre-functional
6.00	-0.39	0.51	478	Level 2	Beginners
7.00	-0.14	0.49	509	Level 2	Beginners
8.00	0.10	0.48	537	Level 2	Beginners
9.00	0.32	0.47	565	Level 2	Beginners
10.00	0.54	0.47	592	Level 2	Beginners
11.00	0.76	0.47	618	Level 3	Intermediate
12.00	0.98	0.48	646	Level 3	Intermediate
13.00	1.22	0.49	674	Level 3	Intermediate
14.00	1.46	0.51	704	Level 4	Advanced
15.00	1.73	0.53	737	Level 4	Advanced
16.00	2.04	0.57	774	Level 4	Advanced
17.00	2.40	0.64	819	Level 4	Advanced
18.00	2.88	0.76	877	Level 5	Full English Proficiency
19.00	3.65	1.03	908	Level 5	Full English Proficiency
20.00	4.41	1.46	926	Level 5	Full English Proficiency

Table C4. Conversion Table — Grades 6–8 Reading Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.48	1.48	151	Level 1	Pre-functional
1.00	-2.70	1.04	197	Level 1	Pre-functional
2.00	-1.91	0.77	292	Level 1	Pre-functional
3.00	-1.42	0.65	353	Level 1	Pre-functional
4.00	-1.04	0.59	399	Level 1	Pre-functional
5.00	-0.72	0.55	438	Level 1	Pre-functional
6.00	-0.44	0.52	473	Level 2	Beginners
7.00	-0.18	0.50	504	Level 2	Beginners
8.00	0.07	0.49	534	Level 2	Beginners
9.00	0.30	0.48	563	Level 2	Beginners
10.00	0.53	0.48	591	Level 2	Beginners
11.00	0.76	0.48	619	Level 3	Intermediate
12.00	1.00	0.49	648	Level 3	Intermediate
13.00	1.24	0.50	677	Level 3	Intermediate
14.00	1.50	0.52	709	Level 4	Advanced
15.00	1.78	0.54	743	Level 4	Advanced
16.00	2.10	0.58	782	Level 4	Advanced
17.00	2.48	0.65	828	Level 4	Advanced
18.00	2.97	0.76	888	Level 5	Full English Proficiency
19.00	3.74	1.04	911	Level 5	Full English Proficiency
20.00	4.52	1.47	928	Level 5	Full English Proficiency

Table C5. Conversion Table — Grades 9–12 Reading Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.02	1.47	162	Level 1	Pre-functional
1.00	-2.25	1.04	251	Level 1	Pre-functional
2.00	-1.48	0.76	345	Level 1	Pre-functional
3.00	-0.99	0.64	404	Level 1	Pre-functional
4.00	-0.62	0.58	450	Level 1	Pre-functional
5.00	-0.31	0.54	487	Level 1	Pre-functional
6.00	-0.04	0.51	520	Level 1	Pre-functional
7.00	0.21	0.49	551	Level 2	Beginners
8.00	0.44	0.48	579	Level 2	Beginners
9.00	0.66	0.47	606	Level 2	Beginners
10.00	0.88	0.47	633	Level 3	Intermediate
11.00	1.10	0.47	660	Level 3	Intermediate
12.00	1.32	0.47	687	Level 3	Intermediate
13.00	1.55	0.49	715	Level 3	Intermediate
14.00	1.80	0.50	745	Level 4	Advanced
15.00	2.06	0.53	777	Level 4	Advanced
16.00	2.37	0.57	814	Level 4	Advanced
17.00	2.73	0.64	859	Level 5	Full English Proficiency
18.00	3.21	0.76	898	Level 5	Full English Proficiency
19.00	3.98	1.03	916	Level 5	Full English Proficiency
20.00	4.74	1.46	933	Level 5	Full English Proficiency

Table C6. Conversion Table — Grades 9–12 Reading Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-2.92	1.46	169	Level 1	Pre-functional
1.00	-2.17	1.03	261	Level 1	Pre-functional
2.00	-1.41	0.75	354	Level 1	Pre-functional
3.00	-0.94	0.63	411	Level 1	Pre-functional
4.00	-0.58	0.57	455	Level 1	Pre-functional
5.00	-0.28	0.53	491	Level 1	Pre-functional
6.00	-0.02	0.50	523	Level 1	Pre-functional
7.00	0.22	0.48	552	Level 2	Beginners
8.00	0.45	0.47	580	Level 2	Beginners
9.00	0.66	0.46	606	Level 2	Beginners
10.00	0.87	0.46	632	Level 3	Intermediate
11.00	1.09	0.46	658	Level 3	Intermediate
12.00	1.30	0.47	685	Level 3	Intermediate
13.00	1.53	0.48	712	Level 3	Intermediate
14.00	1.77	0.50	742	Level 4	Advanced
15.00	2.04	0.53	774	Level 4	Advanced
16.00	2.34	0.57	811	Level 4	Advanced
17.00	2.70	0.64	855	Level 5	Full English Proficiency
18.00	3.18	0.76	898	Level 5	Full English Proficiency
19.00	3.94	1.03	915	Level 5	Full English Proficiency
20.00	4.70	1.46	933	Level 5	Full English Proficiency

Table C7. Conversion Table — Grades 3–5 Listening Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.41	1.47	161	Level 1	Pre-functional
1.00	-2.63	1.04	236	Level 1	Pre-functional
2.00	-1.85	0.77	329	Level 1	Pre-functional
3.00	-1.35	0.65	388	Level 1	Pre-functional
4.00	-0.96	0.59	434	Level 1	Pre-functional
5.00	-0.64	0.55	473	Level 2	Beginners
6.00	-0.35	0.53	507	Level 2	Beginners
7.00	-0.08	0.51	539	Level 2	Beginners
8.00	0.18	0.50	570	Level 3	Intermediate
9.00	0.43	0.50	599	Level 3	Intermediate
10.00	0.68	0.50	629	Level 3	Intermediate
11.00	0.93	0.51	660	Level 4	Advanced
12.00	1.20	0.53	691	Level 4	Advanced
13.00	1.49	0.55	726	Level 5	Full English Proficiency
14.00	1.81	0.59	764	Level 5	Full English Proficiency
15.00	2.19	0.65	810	Level 5	Full English Proficiency
16.00	2.69	0.77	869	Level 5	Full English Proficiency
17.00	3.47	1.04	914	Level 5	Full English Proficiency
18.00	4.25	1.47	928	Level 5	Full English Proficiency

Table C8. Conversion Table — Grades 3–5 Listening Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.38	1.47	162	Level 1	Pre-functional
1.00	-2.61	1.04	238	Level 1	Pre-functional
2.00	-1.83	0.77	331	Level 1	Pre-functional
3.00	-1.34	0.65	390	Level 1	Pre-functional
4.00	-0.96	0.59	435	Level 1	Pre-functional
5.00	-0.64	0.55	473	Level 2	Beginners
6.00	-0.35	0.52	507	Level 2	Beginners
7.00	-0.09	0.51	539	Level 2	Beginners
8.00	0.17	0.50	569	Level 3	Intermediate
9.00	0.41	0.50	598	Level 3	Intermediate
10.00	0.66	0.50	627	Level 3	Intermediate
11.00	0.91	0.51	657	Level 4	Advanced
12.00	1.17	0.52	689	Level 4	Advanced
13.00	1.46	0.55	723	Level 4	Advanced
14.00	1.78	0.59	761	Level 5	Full English Proficiency
15.00	2.16	0.65	806	Level 5	Full English Proficiency
16.00	2.65	0.77	865	Level 5	Full English Proficiency
17.00	3.43	1.04	914	Level 5	Full English Proficiency
18.00	4.21	1.47	928	Level 5	Full English Proficiency

Table C9. Conversion Table — Grades 6–8 Listening Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.00	1.47	191	Level 1	Pre-functional
1.00	-2.23	1.04	284	Level 1	Pre-functional
2.00	-1.45	0.77	376	Level 1	Pre-functional
3.00	-0.95	0.65	435	Level 1	Pre-functional
4.00	-0.57	0.59	480	Level 1	Pre-functional
5.00	-0.25	0.55	519	Level 1	Pre-functional
6.00	0.03	0.52	553	Level 1	Pre-functional
7.00	0.30	0.51	585	Level 2	Beginners
8.00	0.55	0.50	615	Level 2	Beginners
9.00	0.80	0.50	644	Level 3	Intermediate
10.00	1.05	0.50	674	Level 3	Intermediate
11.00	1.30	0.51	704	Level 3	Intermediate
12.00	1.57	0.52	736	Level 4	Advanced
13.00	1.86	0.55	770	Level 4	Advanced
14.00	2.18	0.59	808	Level 5	Full English Proficiency
15.00	2.56	0.65	853	Level 5	Full English Proficiency
16.00	3.05	0.77	907	Level 5	Full English Proficiency
17.00	3.83	1.04	921	Level 5	Full English Proficiency
18.00	4.60	1.47	935	Level 5	Full English Proficiency

Table C10. Conversion Table — Grades 6–8 Listening Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.12	1.48	178	Level 1	Pre-functional
1.00	-2.33	1.05	272	Level 1	Pre-functional
2.00	-1.54	0.77	366	Level 1	Pre-functional
3.00	-1.03	0.66	426	Level 1	Pre-functional
4.00	-0.64	0.60	473	Level 1	Pre-functional
5.00	-0.31	0.56	512	Level 1	Pre-functional
6.00	-0.01	0.53	547	Level 1	Pre-functional
7.00	0.26	0.52	580	Level 2	Beginners
8.00	0.53	0.51	612	Level 2	Beginners
9.00	0.79	0.51	642	Level 3	Intermediate
10.00	1.05	0.51	673	Level 3	Intermediate
11.00	1.31	0.52	705	Level 3	Intermediate
12.00	1.59	0.54	739	Level 4	Advanced
13.00	1.90	0.57	775	Level 4	Advanced
14.00	2.24	0.61	816	Level 5	Full English Proficiency
15.00	2.65	0.67	864	Level 5	Full English Proficiency
16.00	3.17	0.79	909	Level 5	Full English Proficiency
17.00	3.98	1.06	923	Level 5	Full English Proficiency
18.00	4.79	1.49	938	Level 5	Full English Proficiency

Table C11. Conversion Table — Grades 9–12 Listening Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-2.87	1.47	207	Level 1	Pre-functional
1.00	-2.09	1.04	300	Level 1	Pre-functional
2.00	-1.31	0.77	392	Level 1	Pre-functional
3.00	-0.82	0.65	451	Level 1	Pre-functional
4.00	-0.44	0.59	496	Level 1	Pre-functional
5.00	-0.12	0.54	534	Level 1	Pre-functional
6.00	0.15	0.52	567	Level 2	Beginners
7.00	0.41	0.50	598	Level 2	Beginners
8.00	0.65	0.48	626	Level 2	Beginners
9.00	0.88	0.48	654	Level 3	Intermediate
10.00	1.11	0.47	680	Level 3	Intermediate
11.00	1.33	0.48	707	Level 3	Intermediate
12.00	1.56	0.48	734	Level 4	Advanced
13.00	1.80	0.49	763	Level 4	Advanced
14.00	2.05	0.51	792	Level 4	Advanced
15.00	2.32	0.54	825	Level 4	Advanced
16.00	2.63	0.58	862	Level 5	Full English Proficiency
17.00	3.00	0.64	906	Level 5	Full English Proficiency
18.00	3.48	0.76	915	Level 5	Full English Proficiency
19.00	4.25	1.04	928	Level 5	Full English Proficiency
20.00	5.02	1.47	942	Level 5	Full English Proficiency

Table C12. Conversion Table — Grades 9–12 Listening Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-2.74	1.46	222	Level 1	Pre-functional
1.00	-1.98	1.03	313	Level 1	Pre-functional
2.00	-1.22	0.75	403	Level 1	Pre-functional
3.00	-0.75	0.64	460	Level 1	Pre-functional
4.00	-0.39	0.57	503	Level 1	Pre-functional
5.00	-0.09	0.53	538	Level 1	Pre-functional
6.00	0.18	0.50	570	Level 2	Beginners
7.00	0.42	0.48	599	Level 2	Beginners
8.00	0.65	0.47	626	Level 2	Beginners
9.00	0.86	0.46	652	Level 3	Intermediate
10.00	1.08	0.46	677	Level 3	Intermediate
11.00	1.29	0.47	703	Level 3	Intermediate
12.00	1.51	0.47	729	Level 4	Advanced
13.00	1.74	0.48	756	Level 4	Advanced
14.00	1.98	0.50	785	Level 4	Advanced
15.00	2.25	0.53	817	Level 4	Advanced
16.00	2.55	0.57	853	Level 5	Full English Proficiency
17.00	2.92	0.64	896	Level 5	Full English Proficiency
18.00	3.40	0.76	913	Level 5	Full English Proficiency
19.00	4.16	1.03	927	Level 5	Full English Proficiency
20.00	4.92	1.46	940	Level 5	Full English Proficiency

Table C13. Conversion Table — Grades 3–5 Speaking Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-4.05	1.47	205	Level 1	Pre-functional
1.00	-3.28	1.04	276	Level 1	Pre-functional
2.00	-2.51	0.76	347	Level 1	Pre-functional
3.00	-2.04	0.63	390	Level 1	Pre-functional
4.00	-1.68	0.56	423	Level 1	Pre-functional
5.00	-1.39	0.52	450	Level 1	Pre-functional
6.00	-1.14	0.49	473	Level 2	Beginners
7.00	-0.91	0.47	494	Level 2	Beginners
8.00	-0.69	0.46	514	Level 2	Beginners
9.00	-0.49	0.45	533	Level 2	Beginners
10.00	-0.28	0.45	552	Level 3	Intermediate
11.00	-0.09	0.45	570	Level 3	Intermediate
12.00	0.11	0.45	588	Level 3	Intermediate
13.00	0.31	0.45	607	Level 3	Intermediate
14.00	0.52	0.45	625	Level 3	Intermediate
15.00	0.72	0.46	644	Level 3	Intermediate
16.00	0.94	0.47	664	Level 3	Intermediate
17.00	1.17	0.49	685	Level 4	Advanced
18.00	1.41	0.51	708	Level 4	Advanced
19.00	1.68	0.53	733	Level 4	Advanced
20.00	1.99	0.58	761	Level 4	Advanced
21.00	2.36	0.64	795	Level 4	Advanced
22.00	2.85	0.77	840	Level 5	Full English Proficiency
23.00	3.63	1.05	904	Level 5	Full English Proficiency
24.00	4.41	1.48	929	Level 5	Full English Proficiency

Table C14. Conversion Table — Grades 3–5 Speaking Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-4.03	1.46	207	Level 1	Pre-functional
1.00	-3.27	1.03	277	Level 1	Pre-functional
2.00	-2.51	0.75	347	Level 1	Pre-functional
3.00	-2.04	0.63	390	Level 1	Pre-functional
4.00	-1.69	0.56	423	Level 1	Pre-functional
5.00	-1.40	0.51	449	Level 1	Pre-functional
6.00	-1.15	0.48	472	Level 2	Beginners
7.00	-0.93	0.46	492	Level 2	Beginners
8.00	-0.73	0.45	511	Level 2	Beginners
9.00	-0.53	0.44	529	Level 2	Beginners
10.00	-0.34	0.43	546	Level 2	Beginners
11.00	-0.16	0.43	563	Level 3	Intermediate
12.00	0.03	0.44	581	Level 3	Intermediate
13.00	0.22	0.44	598	Level 3	Intermediate
14.00	0.42	0.45	617	Level 3	Intermediate
15.00	0.63	0.46	636	Level 3	Intermediate
16.00	0.85	0.48	656	Level 3	Intermediate
17.00	1.09	0.50	678	Level 4	Advanced
18.00	1.35	0.52	702	Level 4	Advanced
19.00	1.64	0.55	728	Level 4	Advanced
20.00	1.97	0.60	759	Level 4	Advanced
21.00	2.37	0.67	796	Level 4	Advanced
22.00	2.90	0.79	845	Level 5	Full English Proficiency
23.00	3.73	1.07	907	Level 5	Full English Proficiency
24.00	4.56	1.51	934	Level 5	Full English Proficiency

Table C15. Conversion Table — Grades 6–8 Speaking Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.01	1.39	301	Level 1	Pre-functional
1.00	-2.35	0.97	362	Level 1	Pre-functional
2.00	-1.69	0.69	422	Level 1	Pre-functional
3.00	-1.29	0.58	459	Level 2	Beginners
4.00	-1.00	0.52	486	Level 2	Beginners
5.00	-0.75	0.48	509	Level 2	Beginners
6.00	-0.54	0.45	528	Level 2	Beginners
7.00	-0.34	0.43	546	Level 2	Beginners
8.00	-0.16	0.42	563	Level 2	Beginners
9.00	0.01	0.41	579	Level 2	Beginners
10.00	0.17	0.41	594	Level 2	Beginners
11.00	0.34	0.40	609	Level 2	Beginners
12.00	0.50	0.41	624	Level 3	Intermediate
13.00	0.67	0.41	639	Level 3	Intermediate
14.00	0.84	0.42	655	Level 3	Intermediate
15.00	1.02	0.43	671	Level 3	Intermediate
16.00	1.21	0.44	689	Level 3	Intermediate
17.00	1.41	0.47	708	Level 3	Intermediate
18.00	1.64	0.49	729	Level 4	Advanced
19.00	1.91	0.53	753	Level 4	Advanced
20.00	2.22	0.59	782	Level 4	Advanced
21.00	2.61	0.67	818	Level 4	Advanced
22.00	3.14	0.79	866	Level 5	Full English Proficiency
23.00	3.96	1.07	915	Level 5	Full English Proficiency
24.00	4.79	1.51	941	Level 5	Full English Proficiency

Table C16. Conversion Table — Grades 6–8 Speaking Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.13	1.40	290	Level 1	Pre-functional
1.00	-2.45	0.99	352	Level 1	Pre-functional
2.00	-1.77	0.71	415	Level 1	Pre-functional
3.00	-1.36	0.59	453	Level 1	Pre-functional
4.00	-1.05	0.53	481	Level 2	Beginners
5.00	-0.79	0.49	505	Level 2	Beginners
6.00	-0.57	0.46	525	Level 2	Beginners
7.00	-0.37	0.44	544	Level 2	Beginners
8.00	-0.18	0.43	562	Level 2	Beginners
9.00	0.00	0.42	578	Level 2	Beginners
10.00	0.18	0.41	594	Level 2	Beginners
11.00	0.35	0.41	610	Level 2	Beginners
12.00	0.51	0.41	625	Level 3	Intermediate
13.00	0.68	0.41	641	Level 3	Intermediate
14.00	0.86	0.42	657	Level 3	Intermediate
15.00	1.04	0.43	673	Level 3	Intermediate
16.00	1.23	0.44	691	Level 3	Intermediate
17.00	1.43	0.46	710	Level 3	Intermediate
18.00	1.66	0.49	731	Level 4	Advanced
19.00	1.92	0.53	755	Level 4	Advanced
20.00	2.23	0.58	783	Level 4	Advanced
21.00	2.61	0.66	818	Level 4	Advanced
22.00	3.12	0.78	865	Level 5	Full English Proficiency
23.00	3.93	1.06	914	Level 5	Full English Proficiency
24.00	4.73	1.49	940	Level 5	Full English Proficiency

Table C17. Conversion Table — Grades 9–12 Speaking Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.19	1.46	284	Level 1	Pre-functional
1.00	-2.44	1.03	353	Level 1	Pre-functional
2.00	-1.69	0.75	423	Level 1	Pre-functional
3.00	-1.22	0.63	465	Level 1	Pre-functional
4.00	-0.87	0.56	498	Level 1	Pre-functional
5.00	-0.58	0.52	524	Level 1	Pre-functional
6.00	-0.33	0.49	548	Level 1	Pre-functional
7.00	-0.09	0.47	569	Level 1	Pre-functional
8.00	0.12	0.46	589	Level 2	Beginners
9.00	0.33	0.45	608	Level 2	Beginners
10.00	0.52	0.44	626	Level 2	Beginners
11.00	0.72	0.44	644	Level 2	Beginners
12.00	0.91	0.44	662	Level 3	Intermediate
13.00	1.10	0.44	679	Level 3	Intermediate
14.00	1.30	0.44	697	Level 3	Intermediate
15.00	1.49	0.45	715	Level 3	Intermediate
16.00	1.70	0.46	734	Level 3	Intermediate
17.00	1.91	0.47	754	Level 3	Intermediate
18.00	2.14	0.49	775	Level 4	Advanced
19.00	2.39	0.52	798	Level 4	Advanced
20.00	2.68	0.56	825	Level 4	Advanced
21.00	3.04	0.63	857	Level 5	Full English Proficiency
22.00	3.51	0.75	900	Level 5	Full English Proficiency
23.00	4.26	1.03	924	Level 5	Full English Proficiency
24.00	5.02	1.46	949	Level 5	Full English Proficiency

Table C18. Conversion Table — Grades 9–12 Speaking Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-2.67	1.39	332	Level 1	Pre-functional
1.00	-2.00	0.98	394	Level 1	Pre-functional
2.00	-1.33	0.70	455	Level 1	Pre-functional
3.00	-0.93	0.58	492	Level 1	Pre-functional
4.00	-0.63	0.52	520	Level 1	Pre-functional
5.00	-0.38	0.48	543	Level 1	Pre-functional
6.00	-0.16	0.45	563	Level 1	Pre-functional
7.00	0.03	0.44	581	Level 2	Beginners
8.00	0.22	0.42	598	Level 2	Beginners
9.00	0.40	0.42	614	Level 2	Beginners
10.00	0.57	0.41	630	Level 2	Beginners
11.00	0.73	0.41	645	Level 2	Beginners
12.00	0.90	0.40	660	Level 3	Intermediate
13.00	1.06	0.41	675	Level 3	Intermediate
14.00	1.23	0.41	691	Level 3	Intermediate
15.00	1.40	0.41	706	Level 3	Intermediate
16.00	1.57	0.42	722	Level 3	Intermediate
17.00	1.75	0.44	739	Level 3	Intermediate
18.00	1.95	0.46	758	Level 3	Intermediate
19.00	2.18	0.49	778	Level 4	Advanced
20.00	2.44	0.53	802	Level 4	Advanced
21.00	2.76	0.60	831	Level 4	Advanced
22.00	3.19	0.73	871	Level 5	Full English Proficiency
23.00	3.91	1.01	913	Level 5	Full English Proficiency
24.00	4.63	1.43	936	Level 5	Full English Proficiency

Table C19. Conversion Table — Grades 3–5 Writing Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.41	1.41	214	Level 1	Pre-functional
1.00	-2.71	0.99	286	Level 1	Pre-functional
2.00	-2.02	0.72	359	Level 1	Pre-functional
3.00	-1.58	0.62	404	Level 1	Pre-functional
4.00	-1.23	0.57	441	Level 1	Pre-functional
5.00	-0.92	0.54	473	Level 2	Beginners
6.00	-0.63	0.54	503	Level 2	Beginners
7.00	-0.35	0.54	533	Level 2	Beginners
8.00	-0.05	0.54	563	Level 2	Beginners
9.00	0.25	0.55	594	Level 3	Intermediate
10.00	0.56	0.57	627	Level 3	Intermediate
11.00	0.90	0.60	662	Level 3	Intermediate
12.00	1.29	0.64	702	Level 4	Advanced
13.00	1.75	0.72	750	Level 4	Advanced
14.00	2.34	0.83	811	Level 5	Full English Proficiency
15.00	3.16	0.99	894	Level 5	Full English Proficiency
16.00	4.35	1.22	924	Level 5	Full English Proficiency
17.00	5.53	1.71	954	Level 5	Full English Proficiency

Table C20. Conversion Table — Grades 3–5 Writing Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.56	1.46	198	Level 1	Pre-functional
1.00	-2.81	1.03	277	Level 1	Pre-functional
2.00	-2.06	0.75	355	Level 1	Pre-functional
3.00	-1.58	0.64	405	Level 1	Pre-functional
4.00	-1.20	0.59	444	Level 1	Pre-functional
5.00	-0.87	0.56	478	Level 2	Beginners
6.00	-0.57	0.54	509	Level 2	Beginners
7.00	-0.28	0.54	540	Level 2	Beginners
8.00	0.01	0.54	570	Level 2	Beginners
9.00	0.31	0.55	601	Level 3	Intermediate
10.00	0.62	0.57	633	Level 3	Intermediate
11.00	0.96	0.60	668	Level 3	Intermediate
12.00	1.34	0.64	707	Level 4	Advanced
13.00	1.79	0.71	754	Level 4	Advanced
14.00	2.37	0.82	815	Level 5	Full English Proficiency
15.00	3.18	0.98	894	Level 5	Full English Proficiency
16.00	4.35	1.22	924	Level 5	Full English Proficiency
17.00	5.52	1.70	954	Level 5	Full English Proficiency

Table C21. Conversion Table — Grades 6–8 Writing Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.33	1.46	222	Level 1	Pre-functional
1.00	-2.57	1.03	301	Level 1	Pre-functional
2.00	-1.81	0.76	380	Level 1	Pre-functional
3.00	-1.32	0.65	432	Level 1	Pre-functional
4.00	-0.93	0.60	472	Level 1	Pre-functional
5.00	-0.59	0.57	507	Level 1	Pre-functional
6.00	-0.28	0.55	540	Level 1	Pre-functional
7.00	0.03	0.55	571	Level 2	Beginners
8.00	0.33	0.55	602	Level 2	Beginners
9.00	0.63	0.56	634	Level 2	Beginners
10.00	0.95	0.58	668	Level 3	Intermediate
11.00	1.30	0.61	704	Level 3	Intermediate
12.00	1.70	0.65	745	Level 4	Advanced
13.00	2.16	0.71	792	Level 4	Advanced
14.00	2.73	0.81	852	Level 4	Advanced
15.00	3.50	0.97	903	Level 5	Full English Proficiency
16.00	4.71	1.26	933	Level 5	Full English Proficiency
17.00	5.91	1.72	964	Level 5	Full English Proficiency

Table C22. Conversion Table — Grades 6–8 Writing Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.13	1.44	243	Level 1	Pre-functional
1.00	-2.40	1.02	319	Level 1	Pre-functional
2.00	-1.67	0.74	395	Level 1	Pre-functional
3.00	-1.21	0.63	442	Level 1	Pre-functional
4.00	-0.86	0.57	479	Level 1	Pre-functional
5.00	-0.56	0.54	511	Level 1	Pre-functional
6.00	-0.27	0.53	540	Level 1	Pre-functional
7.00	0.00	0.53	569	Level 2	Beginners
8.00	0.28	0.53	598	Level 2	Beginners
9.00	0.57	0.55	628	Level 2	Beginners
10.00	0.88	0.57	660	Level 3	Intermediate
11.00	1.22	0.60	695	Level 3	Intermediate
12.00	1.60	0.64	735	Level 4	Advanced
13.00	2.05	0.70	782	Level 4	Advanced
14.00	2.61	0.80	840	Level 4	Advanced
15.00	3.37	0.94	899	Level 5	Full English Proficiency
16.00	4.48	1.20	927	Level 5	Full English Proficiency
17.00	5.59	1.67	956	Level 5	Full English Proficiency

Table C23. Conversion Table — Grades 9–12 Writing Form A

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-3.15	1.44	241	Level 1	Pre-functional
1.00	-2.42	1.02	317	Level 1	Pre-functional
2.00	-1.69	0.74	393	Level 1	Pre-functional
3.00	-1.24	0.62	440	Level 1	Pre-functional
4.00	-0.89	0.56	476	Level 1	Pre-functional
5.00	-0.60	0.52	506	Level 1	Pre-functional
6.00	-0.34	0.50	533	Level 2	Beginners
7.00	-0.09	0.49	559	Level 2	Beginners
8.00	0.15	0.49	584	Level 2	Beginners
9.00	0.39	0.49	609	Level 2	Beginners
10.00	0.64	0.50	635	Level 3	Intermediate
11.00	0.90	0.52	662	Level 3	Intermediate
12.00	1.18	0.54	692	Level 3	Intermediate
13.00	1.50	0.57	724	Level 4	Advanced
14.00	1.85	0.61	760	Level 4	Advanced
15.00	2.26	0.67	803	Level 4	Advanced
16.00	2.78	0.77	857	Level 5	Full English Proficiency
17.00	3.52	0.97	903	Level 5	Full English Proficiency
18.00	4.81	1.29	936	Level 5	Full English Proficiency
19.00	6.55	1.34	980	Level 5	Full English Proficiency
20.00	8.29	2.02	1024	Level 5	Full English Proficiency

Table C24. Conversion Table — Grades 9–12 Writing Form B

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0.00	-2.90	1.43	268	Level 1	Pre-functional
1.00	-2.19	1.01	341	Level 1	Pre-functional
2.00	-1.48	0.72	415	Level 1	Pre-functional
3.00	-1.04	0.60	460	Level 1	Pre-functional
4.00	-0.72	0.54	494	Level 1	Pre-functional
5.00	-0.45	0.50	522	Level 2	Beginners
6.00	-0.20	0.48	547	Level 2	Beginners
7.00	0.03	0.47	571	Level 2	Beginners
8.00	0.25	0.47	594	Level 2	Beginners
9.00	0.47	0.48	617	Level 2	Beginners
10.00	0.70	0.49	641	Level 3	Intermediate
11.00	0.95	0.50	667	Level 3	Intermediate
12.00	1.21	0.52	694	Level 3	Intermediate
13.00	1.50	0.55	724	Level 4	Advanced
14.00	1.82	0.59	758	Level 4	Advanced
15.00	2.20	0.64	797	Level 4	Advanced
16.00	2.66	0.72	845	Level 4	Advanced
17.00	3.28	0.85	897	Level 5	Full English Proficiency
18.00	4.18	1.06	920	Level 5	Full English Proficiency
19.00	5.62	1.35	956	Level 5	Full English Proficiency
20.00	7.06	1.83	993	Level 5	Full English Proficiency