

Nebraska

**English Language Development Assessment
(ELDA)**

TECHNICAL REPORT

2015 Operational Test Administration

September 2015



423 Morris Street, Durham, North Carolina 27701
(919) 683-2413

www.measurementinc.com

ENGLISH LANGUAGE DEVELOPMENT ASSESSMENT (ELDA)

Developed by

The English Language Development Assessment (ELDA)
State Collaborative on Assessment and Student Standards (SCASS)

With operational support provided by

Measurement Incorporated

The ELDA is a product of the collaboration among ELDA (formerly ELL) SCASS member states, the Council of Chief State School Officers (CCSSO), and the U.S. Department of Education. Funding to support development of this assessment has come from ELL SCASS member states, CCSSO, and the U.S. Department of Education through a Section 6112 Enhanced Assessment Instruments grant.

The Council would like to acknowledge and express its gratitude to the ELL and ELDA SCASS members [current and former] who have dedicated their time and insights to ensure that ELDA not only meets the requirements of the law, but also meets the needs of students and teachers. The current ELDA SCASS member states whose staff continues to contribute significantly to the operation and development of the ELDA include:

Arkansas
Iowa

Louisiana
Nebraska
South Carolina

Tennessee
West Virginia

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Terry Holiday (Kentucky), President

Chris Minnich, Executive Director

Scott Norton, Strategic Initiative Director, Standards, Assessment, & Accountability

Sharon M. Sáez, ELDA Coordinator

State Collaborative on Assessment and Student Standards (SCASS)

Council of Chief State School Officers

One Massachusetts Avenue, NW

Suite 700

Washington, D.C. 20001-1431

Copyright© 2015 by the Council of Chief State School Officers. All rights reserved.

Table of Contents

Executive Summary	1
I. Introduction	2
NCLB Requirements	2
Consortium Member States	2
Overview of ELDA	3
Nebraska Use of ELDA	3
II. ELDA Background	4
ELDA Design and Development	4
ELDA 3–12.....	5
ELDA K–2	10
Validity	15
Performance Level Descriptors.....	17
ELDA Operational Administration Procedures	19
Pre-Administration.....	19
Administration	21
Post-Administration	23
ELDA Scoring and Reporting Procedures	24
Scoring, ELDA K–2.....	24
Scoring, ELDA 3–12	24
Answer document—For Official Use Only	25
Comprehension, Production, and Composite Levels	25
Reporting.....	27
III. Nebraska ELDA 2015 Administration	29
Administration Summary	29
Technical Characteristics	29
Data preparation and quality check.....	29
Item and Test Analysis.....	30
Performance levels.....	32
IV. References	33

Appendices

Appendix A: 2015 Conversion Tables	34
Appendix B: Performance Level Descriptors.....	46
Appendix C: 2015 Sample Score Reports	57
Appendix D: 2015 Operational Item Statistics.....	62
Appendix E: 2015 Raw Score Frequency Tables.....	78
Appendix F: 2015 Operational Test Summary Statistics.....	99
Appendix G: 2015 Operational Raw Score Summary Statistics	101
Appendix H: 2015 Performance Level Classifications	107

List of Tables

Table II.1	ELDA Development Timeline.....	4
Table II.2	Final Student Counts for the 2005 ELDA 3–12 Operational Tests (Form 1).....	8
Table II.3	ELDA 3–12 Recommended Final Raw Cut Scores	9
Table II.4	ELDA 3–12 Short and Original Form Structure.....	10
Table II.5	ELDA K–2 Preliminary Cut Scores.....	13
Table II.6	ELDA K–2 Cut Scores.....	14
Table II.7	ELDA K–2 2006 Operational Administration.....	14
Table II.8	Suggested ELDA Test Administration Times	22
Table II.9	Rules for Combining Listening and Reading.....	26
Table II.10	Rules for Combining Speaking and Writing ELDA K–2	26
Table II.11	Rules for Combining Comprehension and Production ELDA K–2.....	26
Table II.12	ICPL Weights ELDA 3–12.....	27
Table II.13	ICPL Cut Scores ELDA 3–12.....	27
Table III.1	2015 ELDA Scanned Answer Documents.....	29
Table III.2	2015 RT Answer Documents, by Cluster	30

Executive Summary

The English Language Development Assessment (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, also called English language learners (ELLs). The battery consists of four separate tests for Listening, Speaking, Reading, and Writing. In addition, a Comprehension score is derived from a combination of the Listening and Reading scores, and a Composite score is based on a weighted combination of all four language domains. The purpose of this Technical Report is threefold: (1) to provide background information about ELDA development and use; (2) to outline ELDA administration, scoring, and reporting procedures; and (3) to present the technical characteristics of the battery as administered operationally in the spring of 2015 in Nebraska.

In response to the implementation of NCLB in 2002, the Limited English Proficient State Collaborative on Assessment and Student Standards (LEP SCASS) of the Council of Chief State School Officers (CCSSO) led a consortium of states and expert consultants in the design, development, and implementation of ELDA. ELDA for grades 3 through 12 (ELDA 3–12) was consequently developed beginning in 2003, initially field-tested in 2004, and administered operationally for the first time in 2005. ELDA for kindergarten through grade 2 (ELDA K–2) was developed beginning in 2004, initially field-tested in 2005, and administered operationally for the first time in 2006. Nebraska has been a member of the LEP SCASS since the original design and development of ELDA and had significant input into the process. ELLs in Nebraska have participated in all field tests and operational administrations of ELDA since the inception of the assessment.

ELDA K–2 is structured as a series of inventories; test administrators score items throughout the testing window, and completed answer documents are forwarded to Measurement Incorporated (MI) for further processing and reporting. ELDA 3–12 consists of multiple-choice items, student-generated responses to writing prompts, and spoken responses to recorded prompts. Multiple-choice items and writing responses are scored by MI; the former are machine-scanned and scored, while the latter are hand-scored by trained readers. Spoken response items are scored by test administrators in real time and subsequently submitted to MI.

In 2015, as has been the case since 2011, Nebraska administered a shortened ELDA battery during the six-week testing window from Monday, February 9 through Friday, March 20. Materials are provided to testing sites at least one week in advance of the testing window.

I. Introduction

The English Language Development Assessment (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, also called English language learners (ELLs). The battery consists of four separate tests for Listening, Speaking, Reading, and Writing. In addition, a Comprehension score is derived from a combination of the Listening and Reading scores, and a Composite score is based on a weighted combination of all four language domains. The purpose of this Technical Report is threefold: (1) to provide background information about ELDA development and use; (2) to outline ELDA administration, scoring, and reporting procedures; and (3) to present the technical characteristics of the battery as administered operationally in the spring of 2015 in Nebraska. In this introduction, we lay the groundwork for these discussions by providing some historical context.

NCLB Requirements

In 2001, the Elementary and Secondary Education Act was reauthorized as the No Child Left Behind Act of 2001 (NCLB). Among other requirements, NCLB calls for state planning and accountability practices regarding the achievement of students with limited English proficiency (English Language Learners, or ELLs). Specifically, Title I, *Improving the Academic Achievement of the Disadvantaged*, and Title III, *Language Instruction For Limited English Proficient and Immigrant Students*, require that state and local education agencies develop plans to ensure that 100 percent of ELL students achieve the state’s “proficient” level by the 2013–2014 school year, and that these plans include “separate, measurable objectives for continuous and substantial improvement” for students with limited English proficiency.

In response to the implementation of NCLB in 2002, the Limited English Proficient State Collaborative on Assessment and Student Standards (LEP SCASS) of the Council of Chief State School Officers (CCSSO) led a consortium of states and expert consultants in the design, development, and implementation of ELDA. Specifically, the use of ELDA enables states to meet the following six NCLB Title III requirements:

- Provide an annual assessment of English proficiency and show annual progress in acquisition.
- Include all LEP students in English language proficiency (ELP) assessments.
- Obtain independent measures of listening, speaking, reading, and writing.
- Report a separate measure for comprehension.
- Measure proficiency in academic language and in the language of social interaction.
- Align assessments with state ELP standards.

Consortium Member States

ELDA was developed beginning in 2003 by the CCSSO along with a consortium of 15 states that were then members of the LEP-SCASS. Thirteen of these states participated in the original field-testing of ELDA 3–12 in 2004; these states included Alabama, Georgia, Iowa, Indiana, Kentucky, Louisiana, Nebraska, New Jersey, Ohio, Oklahoma, South Carolina, Virginia, and West Virginia.

In the fall of 2006, CCSO formed a new SCASS composed of a subset of states from the LEP (now English Language Learners, ELL) SCASS. These states were those that had formally adopted ELDA as their ELP assessment. During 2007–2008, the ELDA SCASS membership included seven states (Arkansas, Iowa, Louisiana, Nebraska, South Carolina, Nebraska, and West Virginia), each of which administered ELDA as its ELP progress measure.

Overview of ELDA

ELDA tests are aligned with the ESL standards of project member states and were developed to provide content coverage across three academic topic areas (English/Language Arts—ELA; Math, 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 many others). (The shortened version of ELDA Speaking does not include School-Environmental content.) They are tests of language skills with content drawn from age-appropriate school curricular and noncurricular sources. They are not tests of academic content (that is, no external or prior content-related knowledge is required to respond to test questions), nor is performance on production skills tests scored in terms of content validity of response beyond what may be supplied in test input.

While the main function of the ELDA tests is to measure annual progress in English language acquisition, they also permit the identification of students who have reached full English proficiency (FEP) or LEP-exit level (that is, a level considered appropriate for successful functioning within the school system at the appropriate grade level). An FEP classification is not intended to imply skills equivalent to those of a native English speaker. The tests are not designed to provide placement information relative to English language courses or programs offered at a school, nor are they designed to provide diagnostic feedback to students and their language teachers.

Nebraska Use of ELDA

Nebraska's participation in all phases of ELDA development has proved crucial to the assessment's evolution and implementation. The state has administered ELDA in some form, either operationally or as a field test, every year since 2004. Nebraska was a member of the LEP SCASS during the initial development of ELDA 3–12 in 2003. Along with twelve other states, Nebraska participated in the original ELDA 3–12 field test in 2004. In 2005, Nebraska adopted the operational ELDA 3–12 to assess the English language acquisition progress of ELL students in the state and also participated in the original ELDA K–2 field test. In 2006, Nebraska joined the (then) newly formed CCSO ELDA SCASS along with six other states and participated in the first operational administration of ELDA K–2. Since then, Nebraska has administered the operational ELDA to all its ELL students in grades K–12.

In 2011, the ELDA SCASS was merged with the CCSO ELL SCASS, and now works with other member states to continue to focus on the assessment of ELLs' English language proficiency and inclusion of ELLs in academic content assessments. The ELDA SCASS is a technical working group that is part of the ELL SCASS.

II. ELDA Background

ELDA Design and Development

In 2001, the Elementary and Secondary Education Act was reauthorized as the No Child Left Behind Act of 2001 (NCLB). Among other requirements, NCLB calls for state planning and accountability practices regarding the achievement of students with limited English proficiency (English Language Learners, or ELLs). Specifically, Title I, *Improving the Academic Achievement of the Disadvantaged*, and Title III, *Language Instruction For Limited English Proficient and Immigrant Students*, require that state and local education agencies develop plans to ensure that 100 percent of ELL students achieve the state’s “proficient” level by the 2013–2014 school year and that these plans include “separate, measurable objectives for continuous and substantial improvement” for LEP students.

Table II.1 ELDA Development Timeline	
Year	Development milestone
2002	CCSSO and ELDA Steering Committee project start-up meeting
2003	ELDA 3–12 item writing
2004	ELDA 3–12 field test (forms A & B) administered in Alabama, Georgia, Iowa, Indiana, Kentucky, Louisiana, Nebraska, New Jersey, Ohio, Oklahoma, South Carolina, Virginia, and West Virginia
2004	ELDA 3–12 field-test calibration
2005	ELDA 3–12 1st operational test administered in Iowa, Louisiana, Nebraska, Ohio, South Carolina, and West Virginia
2005	ELDA 3–12 recalibration
2005	ELDA 3–12 standard setting
2004	ELDA K–2 development plan
2005	ELDA K–2 item writing
2005	ELDA K–2 field test administered in Indiana, Kentucky, Nebraska, New Jersey, Oklahoma, and West Virginia
2005	ELDA K–2 preliminary standard setting
2005	ELDA K–2 teacher evaluation
2006	ELDA K–2 follow-up standard setting
2006	ELDA K–2 1st operational test administered in Nebraska, Ohio, South Carolina, and West Virginia
2006	ELDA K–2 calibration

In response to the implementation of NCLB in 2002, the LEP SCASS led a consortium of states and expert consultants in the design, development, and implementation of ELDA. Specifically, the use of ELDA enables states to meet the following six NCLB Title III requirements:

- Provide an annual assessment of English proficiency and show annual progress in acquisition.
- Include all LEP students in English language proficiency assessments.
- Obtain independent measures of listening, speaking, reading, and writing.
- Report a separate measure for comprehension.

- Measure proficiency in academic language and in the language of social interaction.
- Align assessments with state ELP standards.

ELDA was developed beginning in 2003 by the CCSSO along with a consortium of 15 states that were then members of the LEP-SCASS. Thirteen of these states participated in the original field-testing of ELDA 3–12 in 2004; these states included Alabama, Georgia, Iowa, Indiana, Kentucky, Louisiana, Nebraska, New Jersey, Ohio, Oklahoma, South Carolina, Virginia, and West Virginia.

In the fall of 2006, CCSSO formed a new SCASS composed of a subset of states from the LEP (now English Language Learners, ELL) SCASS. These states were those that had formally adopted ELDA as their ELP assessment. During 2007–2008, the ELDA SCASS membership comprised seven states (one being Nebraska), each of which administered ELDA as its ELP progress measure. Current members are Arkansas, Iowa, Louisiana, Nebraska, South Carolina, Tennessee, and West Virginia.

ELDA 3–12. The design, development, and implementation of ELDA 3–12 was headed by the CCSSO LEP-SCASS. American Institutes for Research (AIR) developed the test items and forms; the Center for the Study of Assessment, Validity, and Evaluation (C-SAVE), based at the University of Maryland, provided validity and reliability research to the project.

The ELDA assessment for grades 3–12 was aligned to state ESL standards through an analysis of the ESL standards of consortium states available to the project at the outset. From an analysis of state ESL standards for each of the four skills domains, AIR constructed and the LEP-SCASS approved a set of core ESL standards which formed the basis for item design. The content that forms the basis for test items in ELDA is distributed across four topic areas. Approximately 75 percent of the items use language from the three curriculum domains of mathematics, science, and technology; English language arts; and social studies. The remaining 25 percent of the items use the social language of interaction among students, teachers, other school personnel, and parents related to school issues.

AIR proposed and the LEP-SCASS approved the following test item specifications for each of the four skill domain areas.

- **Listening** consists of only multiple-choice items. Students listen to five types of text read by a narrator and actors (short phrases, short dialogues, extended dialogues, short presentations, and extended presentations) and answer comprehension questions; Cluster 3 (grades 3–5) excludes extended presentations because of developmental inappropriateness.
- **Reading** is also entirely composed of multiple-choice items. Students read three types of text (short, early reading comprehension passages or cloze items; instructions; and long passages) and answer comprehension questions.
- **Writing** comprises both multiple-choice and constructed-response items (short and extended). 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 appear in sets of four prompts, each eliciting a different speaking function.

ELDA English Language Proficiency Standards and Test and Item Specifications for Grades 3–12 (AIR, 2005) and *Aligning English Language Proficiency Tests to English Language Learning Standards* (Cook, 2005) together provide a comprehensive discussion of ELDA standards and specifications for grades 3–12.

Item development and review. To develop items that measure these academic standards as specified by the content specifications, AIR brought together a mixture of external item writers, NAEP foreign language item writers, and other internal content experts. The LEP-SCASS member states recommended teachers to AIR who had experience with assessment development, and AIR contacted those teachers and selected them based on their availabilities. Bill Eilfort and Natalie Chen, assessment development consultants, worked to guide item writers during a weekend workshop in Denver in February 2003. The consultants, along with AIR staff, working in groups by domain and grade level, trained the item writers by explaining general item-writing principles and were available to help the participants develop their items. AIR also provided books and texts and other reference materials for the item writers, who subsequently developed items individually and in small groups. After items were drafted and reviewed by the writers, they were entered into the review protocol as part of AIR’s Item Tracking System (ITS) database. The following review levels were then conducted:

- Preliminary review by junior staff for formatting and basic item construction principles
- LABS review by a trained and certified language accessibility, bias, and sensitivity (LABS) reviewer
- Editorial review for grammar, writing conventions, and clarity
- Senior ESL or English language arts content expert review for match to standards and for measurement integrity

Items that passed all reviews were brought to LEP-SCASS meetings for review, comments, revision, and approval. At SCASS content review meetings, members split into grade-cluster groups and were instructed on the specifications for the items, the standards and the benchmarks, and individually reviewed the items before meeting as a group to accept, revise, reject, or recommend revisions for resubmission of all the items. Those items that survived the final review entered the field-test item pool.

Field testing. In March 2004, a field test of ELDA 3–12 was administered to ELL students in 13 consortium member states: Alabama, Georgia, Iowa, Indiana, Kentucky, Louisiana, Nebraska, New Jersey, Ohio, Oklahoma, South Carolina, Virginia, and West Virginia. The purpose of this field test was to ensure the development of an operational form for implementation during the 2004–2005 academic year by using a multistage review grounded in commonly accepted content and psychometric standards.

AIR developed and MI implemented a field-test sampling plan for the 2004 field-test administration. A target of 1,000 students per form was set to obtain reliable estimates for item screening and parameter estimation. The number of students requested from each state was determined by the number of states participating in the field test. A sample of 2,000 students per grade cluster was drawn equally from the thirteen participating member states.

AIR collected data on items from two field-test forms (A & B) across four skill domains and three grade clusters. Field-test forms were constructed according to test specifications developed by the Steering Committee, in collaboration with AIR and members of the LEP-SCASS. The 2004 field-test assessments had the following components:

- *Skill Domains:* Listening, Reading, Speaking, Writing
- *Grade Clusters:* Grades 3–5, Grades 6–8, Grades 9–12
- *Language Groups:* LEP Spanish, LEP Other, LEP Exited, Native English Speakers

Operational forms construction. Prior to item review, field-test data were analyzed. Classical item analysis and differential item functioning (DIF) analyses were conducted to detect any potential test administration or scoring problems; test difficulty, biserial correlation, and omit rate were calculated for each test form. Items then passed through a two-stage review process consisting of reviews by a team of AIR psychometricians and by a Joint Committee formed by members of AIR, the CCSSO, MI, and the LEP-SCASS state membership. Recommendations were then made to include items in the operational pool, revise and resubmit items for administration in future field tests, or reject items from consideration. Only those items that passed all stages of review were included in the master item pool for operational form construction.

AIR used Masters' (1982) partial credit model to estimate ELDA item parameters. To implement randomly equivalent groups design, the Item Response Theory (IRT) calibrations were conducted for each field test by setting the mean population ability to zero for each form in each grade cluster. Goodness-of-fit indices were also used to further analyze appropriateness of items. Through IRT calibration of the same 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. To ensure the quality of the linking item pool, AIR used a stepwise deletion procedure when computing the linking constant.

Operational form construction was conducted jointly with AIR psychometricians, AIR content experts, and LEP-SCASS state member representatives. Form construction used the item locations from the Rasch/IRT analyses to pre-equate operational forms. This process resulted in the creation of an operational form for use in the 2004–2005 school year.

For a detailed discussion of the ELDA 3–12 2004 field-test administration and subsequent analysis, please see the *ELDA Technical Report: 2004 Field Test Administration* (AIR and MI, 2005).

Operational administration. In 2005, the ELDA tests were administered operationally for the first time in six states (Iowa, Louisiana, Nebraska, Ohio, South Carolina, and West Virginia). The first ELDA operational form (Form 1) was developed from the 2004 ELDA field-test items and included four subjects (Reading, Writing, Listening, and Speaking) covering three grade clusters (3–5, 6–8, and 9–12). The final student counts for each ELDA test are summarized in Table II.2.

Table II.2 Final Student Counts for the 2005 ELDA 3–12 Operational Tests (Form 1)	
Grade Cluster	Number
3–5	16,813
6–8	11,334
9–12	9,753
<i>Total</i>	<i>37,900</i>

Scaling and calibration. As described earlier, AIR used IRT calibration of the common items embedded in the 2004 ELDA field-test forms to link the measurement scale of the three grade clusters into one scale. Likewise, the 2005 ELDA 3–12 Operational Form 1 included a set of items common to adjacent grade clusters such that the cluster 3 (grades 3–5), cluster 4 (grades 6–8), and cluster 5 (grades 9–12) forms were jointly calibrated for each subject. Winsteps (Linacre, 2008), a Rasch calibration software program, was used in both instances of calibration. Results of a subsequent scale stability analysis, however, indicated a substantial shift in item function for more than 20% of the items on each subject area test. Consequently, the ELDA Technical Advisory Committee (TAC) recommended that ELDA operational Form 1 be scaled based on the 2005 operational administration. ELDA Form 1 now serves as a reference point, such that all subsequent ELDA test forms are linked to these scales. Appendix A provides the raw-to-theta-to-scale conversion tables for the 2013 ELDA tests.

Standard setting. Following the first operational administration of the ELDA in 2005, MI worked with CCSSO staff and advisors from participating states to conduct a three-day standard-setting session in Durham, North Carolina. From August 15 to 19, 2005, panelists from eight ELDA SCASS states met to recommend cut scores for the various levels and forms. The procedures they followed and the results of their deliberations are detailed in the *Final Report on ELDA Standard Setting* (Bunch, 2006). Table II.3 presents the final raw cut scores recommended by the articulation committee.

**Table II.3
ELDA 3–12 Recommended Final Raw Cut Scores**

Test		Level 2 Beginning	Level 3 Intermediate	Level 4 Advanced	Level 5 Fully English Proficient
3-5	Listening (max = 50)	22	31	39	44
	Reading (max = 50)	23	34	38	45
	Writing (max = 25)	9	16	20	25
	Speaking (max = 32)	11	17	24	29
6-8	Listening (max = 50)	26	32	39	46
	Reading (max = 50)	21	33	38	45
	Writing (max = 31)	16	22	25	30
	Speaking (max = 32)	10	20	26	30
9-12	Listening (max = 60)	27	36	45	53
	Reading (max = 60)	27	36	45	54
	Writing (max = 34)	14	22	27	32
	Speaking (max = 32)	14	20	26	29

Development of shortened test form. The time necessary to administer (and score) ELDA has been perceived as excessive by many local districts, and indeed, other consortium-developed tests (i.e., ACCESS-ELLs, CELLA, Mountain West Trust) are generally shorter. In 2009, CCSSO commissioned a series of four studies, using data from several participating states, to gather information for reducing ELDA administration time (Haladyna, 2009). The results of these studies supported the development of shortened, yet still highly reliable, ELDA 3–12 test forms (see Appendix F for 2015 operational reliability information). Consequently, MI developed a set of shortened ELDA 3–12 test forms, which have been used in Nebraska since 2011. Table II.4 depicts the structure of the shortened forms juxtaposed with that of the original forms.

Table II.4
ELDA 3–12 Numbers of Items and Suggested Administration Time
Short Forms vs. Original Forms

Test	Grade Cluster	Number of MC Items		Number of CR Items		Suggested Time (minutes)	
		Short	Original	Short	Original	Short	Original
Listening	3–5	35	50	-	-	35	50
	6–8	35	50	-	-	42	60
	9–12	35	60	-	-	41	70
Reading	3–5	35	50	-	-	42	60
	6–8	35	50	-	-	42	60
	9–12	35	60	-	-	44	70
Speaking	3–5	-	-	12	16	20	25
	6–8	-	-	12	16	20	25
	9–12	-	-	12	16	20	25
Writing	3–5	12–15	12–15 ^a	4	4	60	60
	6–8	15–18	15–18 ^b	4	4	60	60
	9–12	15–18	15–18 ^c	5	5	60	60

^aThe numbers of multiple-choice writing items differ across original ELDA forms. On Form 1, the numbers of MC items for grades 3–5, 6–8, and 9–12 are 12, 18, and 18, respectively; on Forms 2 and 3, there are 15 items for all grades.

^bSee note a above.

^cSee note a above.

ELDA K–2. In 2004, the CCSSO and consortium member states teamed with MI to develop the ELDA assessment for kindergarten through Grade 2. The framework for ELDA K–2 is based on that of the tests for grades 3–12. Development of the K–2 assessments was guided by Standards and Benchmarks developed by AIR, consortium members, and consultants, as well as by general guidelines offered by LEP-SCASS and Early Childhood Committee members. Specifically, MI designed instruments that:

- address reporting requirements of NCLB Titles I and III;
- focus on academic language rather than academic content;
- embed items from each content from 3–5 cluster tests;
- reflect developmental states of K–2 students;
- resemble instructional activities as much as possible; and
- take no more than about 15 minutes per form.

MI representatives met with CCSSO staff and consultants to develop an overall plan for the K–2 assessments. These meetings took place in Austin, Texas (December, 2004), and Orlando, Florida (January, 2005). In addition, MI worked with consultants identified by CCSSO to guide the development of the assessments. MI commissioned the drafting of a concept paper for review by the ELDA K–2 committee at its January meeting in Orlando. That concept paper and associated comments from the committee formed the basis for the K–2 inventories.

Item development and review. In February, 2005, MI invited nine classroom teachers to participate in an item-development session in Durham, North Carolina. Working from the concept paper drafted by Dr. Espinoza, the draft recommendation position paper from NAEYC, and SCASS committee member input, the teachers created inventory entries for Listening, Reading, Speaking, and Writing. The work was collaborative and had as its goal the generation of enough teacher observations/items to construct three field-test inventories.

MI adopted the inventory approach over the traditional multiple-choice (MC) and constructed-response (CR) item approach because of the age and developmental stage of the student population (kindergarten through second grade). Each “item” in the initial inventories was a statement regarding an observable student behavior, such as the following:

- Follows a two-step instruction in a nonacademic setting (e.g., going to the lunchroom)
- Identifies a picture of an object with the same ending sound as ‘cat’
- Uses correct English words (content-, age-, and grade-appropriate) for manipulatives

The item writers used the lists of standards and benchmarks collected from the LEP-SCASS member states in the consortium as their guides, along with other materials they brought to the session and resources supplied by MI.

The final step in item development was the selection of anchor items from the existing tests for grades 3–5. These items were selected on the basis of their relevance to the K–2 content objectives and difficulty indices derived in the summer of 2004. Those items deemed suitable for administration to first and second graders were identified and added to the inventories developed by the item-development committee. The inclusion of anchor items on the K–2 inventories ensured that overall test difficulty remains comparable and cut scores retain meaning over time.

Field testing. Upon approval of the individual entries, MI assembled a single field-test form for administration in the spring of 2005. The form covered all four content areas and included a sample of the selected anchor items from the current assessments for grades 3–5. Two versions were prepared: one for kindergarten only (all entries except the anchor items from grades 3–5) and another for grades 1–2 (all entries, including the grade 3–5 anchor items). Upon approval of the formatted field-test forms, MI included them in the 2005 field test for grades K–12. MI staff printed, shipped, and collected field-test materials; barcoded all secure materials for secure delivery and collection; checked each incoming box for completeness against the outgoing packing list; and noted and followed up on any discrepancies. Because the student booklets were nonscannable, all data in the booklets were hand-entered, using double entry to guarantee 100 percent accuracy of entered data. Data entry operators entered the scores for each inventory, using software that prompted the operator to enter student name, demographic information, and school information, and then prompted them from row to row and automatically summed the row scores to derive a total score. This software also locked out invalid (out-of-range) entries, further enhancing the accuracy of the data. MI scoring leaders conducted training for scorers as they did in 2004, and those scorers evaluated student responses to the writing prompts as well as responses to the speaking prompts, entering scores by barcode number and forwarding them on monitor sheets to information technology (IT) staff. As scorers completed their task, IT staff merged data from the student booklet with score data and the Student Background Questionnaire

(SBQ) to produce a final data set for analysis. A full discussion of the field-test results can be found in the *ELDA K–2 Technical Manual, Spring 2006* (CCSSO, 2005).

Operational preparation. Subsequent to the 2005 field test and associated analyses, there were two key meetings concerning ELDA K–2. The first was in Savannah, Georgia, in July, 2005. At this meeting, state representatives presented many of the concerns voiced by K–2 teachers regarding the length and difficulty of administering the inventories. In August, Dr. Dina Castro joined the MI team of developers to begin revising the inventories with two key goals: shorten the inventories, and provide more support for teachers who administer them. In December, 2005, member state representatives met again in Washington, D.C., to review revised materials. These materials were ultimately approved with modifications during December 2005 and January 2006. The final materials were submitted and approved on January 31, 2006, and were administered to students in the spring of 2006.

Standard setting. Initial standard setting for the ELDA K–2 inventories was based on teacher ratings of student proficiency. Each teacher administering the K–2 field-test inventories in 2005 was asked to rate the proficiency of each student on a scale of 1 to 5 (5 being Fully English Proficient) on each component. The standard application of this procedure (cf. Jaeger, 1989) is to compile the raw scores for all examinees classified by their teachers in Level 1, Level 2, and so on. Cut scores can then be derived in a number of ways. The simplest is to find the midpoint between two means or medians. (For example, if the mean raw score for all students classified by their teachers as Level 1 were 20, and the mean raw score for all students classified by their teachers as Level 2 were 30, the cut score for entry into Level 2 would be 25, the midpoint between 20 and 30.) Table II.5 summarizes the contrasting groups standard setting results.

Subsequent to the ELDA K–2 field test and initial standard setting, educators in ELDA consortium states recommended several changes to the inventories due to concern regarding the length and difficulty of administering the inventories. Following a generalizability study, the results of which confirmed that the inventories could be shortened without a loss in reliability, the inventories were revised with two key goals: shorten the inventories, and provide more support for teachers who administer them.

The changes, however, proved extensive enough to render the 2005 field-test data inapplicable to the standard-setting process; thus, it became necessary to set standards for the revised inventories in Listening, Reading, Speaking, and Writing (one set for kindergarten, another set for grades 1–2) prior to the 2006 administration. These performance standards had to be set without the benefit of operational or field-test data.

Table II.5
ELDA K–2 Preliminary Cut Scores
(Groups based on teacher rating of overall proficiency)

Level	N	Mean	Preliminary Cut
Reading (87 Points)			
1 - Pre-functional	211	23.0	
2 - Beginning	367	44.2	33.6
3 - Intermediate	353	53.7	48.9
4 - Advanced	298	67.3	60.5
5 - Fully English Proficient (FEP)	136	69.7	68.5
<i>Total</i>	<i>1,365</i>		
Listening (21 Points)			
1 - Pre-functional	135	6.8	
2 - Beginning	298	10.5	8.6
3 - Intermediate	533	14.1	12.3
4 - Advanced	471	16.9	15.5
5 - Fully English Proficient (FEP)	268	18.3	17.6
<i>Total</i>	<i>1,705</i>		
Writing (48 Points)			
1 - Pre-functional	171	21.5	
2 - Beginning	335	31.9	26.7
3 - Intermediate	313	37.2	34.5
4 - Advanced	231	41.3	39.2
5 - Fully English Proficient (FEP)	89	40.0	40.6
<i>Total</i>	<i>1,139</i>		
Speaking (36 Points)			
1 - Pre-functional	27	10.9	
2 - Beginning	62	21.7	16.3
3 - Intermediate	149	26.1	23.9
4 - Advanced	137	30.6	28.3
5 - Fully English Proficient (FEP)	119	32.6	31.6
<i>Total</i>	<i>494</i>		

MI conducted a second standard setting in January 2006 using a generalized holistic procedure. In a generalized holistic procedure, the task of panelists is to examine student work samples and classify each one as belonging to one of the five performance levels. As the inventories had yet to be administered operationally, MI staff prepared 20–24 virtual inventories (i.e., inventories with synthetic scores) with scores ranging from near 0 to the maximum possible. ELL educators and other stakeholders in five participating states (Iowa, Louisiana, Nebraska, Ohio, and South Carolina) convened twice by web conference—first for training in the procedure to be used, and later to discuss results of the first round of rating of profiles. A total of 16 individuals participated (9 for the Kindergarten inventories and 7 for the Grades 1–2 inventories). Results are summarized in Table II.6.

Table II.6 ELDA K–2 Cut Scores					
	Test	Level 2 Beginning	Level 3 Intermediate	Level 4 Advanced	Level 5 Fully English Proficient
K	Reading (max = 42)	8	20	36	40
	Listening (max = 21)	4	9	16	19
	Writing (max = 27)	7	16	21	26
	Speaking (max = 24)	6	12	18	22
1–2	Reading (max = 42)	10	22	31	39
	Listening (max = 21)	6	11	16	19
	Writing (max = 27)	8	17	21	26
	Speaking (max = 24)	8	13	18	22

Operational administration. The first operational administration of the ELDA K–2 inventories took place in the spring of 2006. Over 20,000 students from four states—Nebraska, Ohio, South Carolina, and West Virginia—participated in the assessment, as illustrated in Table II.7.

Table II.7 Final Student Counts for the 2006 ELDA K–2 Operational Administration	
Grade Cluster	Number
K	7,753
1–2	12,957
<i>Total</i>	<i>20,710</i>

Scaling and calibration. In 2006, MI conducted a calibration of the ELDA K–2 to create a link to the existing ELDA 3–12 instruments. MI used a common item design whereby anchor items were chosen from the existing assessments to anchor the grade 1–2 tests to the grade 3–5 tests. Items deemed not significantly altered were also used to create a link between the grade 1–2 tests and the kindergarten test. While the link between grades 1–2 and 3–5 was technically successful, the resultant score scale yielded scores in the extremes that could not be reconciled with existing score scales for the upper grades. Therefore, members of the ELDA TAC directed MI staff to make the ELDA K–2 scaling independent of the scaling for grades 3–12. A complete discussion of the vertical scaling process and results can be found in the *Supplement to the ELDA K–2 Technical Report* (Bunch & Joldersma, 2006).

Validity. An important factor in the design and development of ELDA from the outset was the inclusion of empirical validity studies with a focus on the degree to which the assessment measured proficiency in academic English deemed necessary for success in school. The details of validity studies undertaken with respect both to ELDA K–2 and ELDA 3–12 are described in the original ELDA development reports (e.g., *ELDA K–2 Technical Report: Spring 2006* and *ELDA 2004 Technical Report: Field Test Administration*). Furthermore, an independent review of a number of English language proficiency measures (Wolf et. al., 2008) concluded that ELDA developers successfully established the validity of the instrument with rigorously applied research. Included here is a summary of the evidence supporting ELDA validity.

ELDA K–2. Messick (1989) defined validity as an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment (p. 13, emphasis in original). To be valid for this purpose, tests must: 1) measure what is taught in schools, 2) provide a standard that is consistent with what students need to know and be able to do, 3) provide for consistent measurement of that standard over time, and 4) be free of bias. In addition, the tests should demonstrate a reasonable degree of association with other measures of school achievement. The validity of the tests, therefore, is to be established over a period of time through the accumulation of a body of evidence that clearly demonstrates that all these criteria have been met. This same point of view is also reflected in the most recent version of *Standards for Educational and Psychological Testing*, jointly published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999), and followed by most test developers.

- ELDA K–2 inventories are based on a collection of English language acquisition content standards developed by the states participating in the consortium, with the cooperation of national consultants and test development contractors, and in a manner consistent with the latest research on language acquisition in young children.
- At each stage of the development process, classroom teachers and other stakeholders familiar with the student population and content reviewed blueprints, test items (inventory entries), support materials, and manuals.
- As the inventories were field tested, teachers administering them were given opportunities to provide feedback on the relevance of the inventories to their classroom instruction. Based on these evaluations, most teachers expressed satisfaction with the degree of fit.
- Likewise, performance on ELDA K–2 inventories correlates well with teacher judgments of student performance.
- The performance level descriptors (PLDs) represent a collective effort of CCSSO, the participating states, and content expert consultants. These PLDs are very similar to PLDs submitted to classroom teachers who participated in the 2005 field tests. Those teachers endorsed the PLDs by applying them and sorting their students into five performance levels in a way that was generally consistent with the performance of students on the inventories.
- The content standards are well defined and clearly aligned with test items and PLDs.
- The blueprints and test construction guidelines are well established.
- As described earlier in this report, ELDA K–2 inventories were linked to ELDA 3–12 tests and calibrated to ensure that overall test difficulty remains comparable and cut scores retain meaning over time.

- At each stage of development, bias review was an integral part of overall item and test review. Any item or inventory entry suspected of bias was revised or rejected. Indeed, in a recent review of ELP assessments, Wolf and colleagues concluded that attention to potential item bias and fairness throughout ELDA development was particularly noteworthy and rigorous (2008, January).

ELDA 3–12. CCSSO worked with an independent research group, the Center for the Study of Assessment Validity and Evaluation at the University of Maryland, to design and analyze two sets of studies investigating the validity of the assessment. The first set of studies was conducted during a pilot test of the initial items developed for the ELDA; it involved 320 students in 31 schools across 12 states. The second set of studies was conducted during the large-scale field test (approximately 1,000 students per item) of the ELDA in 2004. The first of these studies used expert review of items, teacher judgments of students, and a latent class analysis of student performance to determine whether the items were consistently and reasonably differentiating levels of proficiency as defined by the five ELDA proficiency levels. The second focused on the relationship of ELDA domain scores to teacher judgments of proficiency and to student scores on two other standardized English language proficiency tests. The third reviewed the structure of the test.

Taken together, these studies provide support for using ELDA to appropriately measure the academic English language proficiency of ELLs in grades 3–12; results indicated the following:

- Students understood test administration procedures and were able to give their best performances in all four skill domains.
- Test score reliabilities ranged from 0.77 to 0.92, similar to score reliabilities achieved in state assessment programs.
- ELDA measures relevant, ever-increasing, complex skill sets in the four domains.
- ELDA better differentiates mid-range and more sophisticated language proficiency by domain than other tests, supporting the validity of inferences about students' academic language proficiency stemming from the ELDA.
- The language skills assessed are for the most part cumulative—more complex skills build on simpler skills for most language constructs.
- ELDA functions similarly for students who are in different types of programs, for students with different language backgrounds, and for students at different grade levels within a grade cluster, supporting the idea that the ELDA measures language proficiency consistently for important subgroups of English language learners.

Performance Level Descriptors. Initially, AIR developed performance level descriptors (PLDs) for each of the four language skills tested in the ELDA 3–12 battery—listening, speaking, reading, and writing—as well as for a fifth score for comprehension, derived from a combination of the listening and reading scores as per federal requirements. In December 2005, CCSSO prepared updated versions of PLDs for all five levels of each inventory included in the ELDA K–2 assessment. The levels are Pre-functional (1), Beginning (2), Intermediate (3), Advanced (4), and Fully English Proficient (5). The PLDs underwent a series of review and revision procedures with both ELDA project staff and nonproject staff and were a critical part of ELDA standard setting panelist training. Copies of the updated PLDs are included in Appendix B.

Forms of PLDs. The PLDs exist in two forms, reflecting the two different functions that PLDs serve. One form is a matrix display of the descriptions of performance levels intended for standard-setting purposes. In this format the cell-by-cell display of information facilitates the task of capturing the distinctions between levels for each of the performance indicators that make up the language domain. The term “indicators” refers to the set of component skills that defines each of the language domains.

The second form is a narrative of the information contained in the matrix, intended for reporting purposes. In this format, the description captures the character of a level across all the indicators, thus making it easier for stakeholders, such as schools and parents, to interpret test performance and progress in the acquisition of English. This document first presents the matrix form of the PLDs and then the narratives. The PLDs are common across the grade clusters for which the tests have been designed. The PLD for listening, for example, provides common performance level descriptions for the 3–5, 6–8, and 9–12 clusters. Underlying the notion of a common performance scale to describe English language proficiency development across 3–12 is the assumption that the same domain performance indicators, as well as the values given to each indicator at each level, used to describe language development in third grade can also be used in high school. What *is* variable across age or grade, with respect to test design and performance measurement, are specific age-appropriate test input materials (contexts for language use such as stimuli, the topics embedded in stimuli, specific features of grammar and vocabulary that are bound by cognitive development, and test graphics) and the cognitive skills required by the language tasks of the test.

Numbers of levels and range. The Steering Committee, having met at project start-up in Berkeley, California, in December 2002, determined that the PLDs should contain five levels of discrimination. Five levels were considered appropriate to (a) capture the construct of development in academic English language proficiency from *Pre-functional* to *Full English Proficiency*; (b) permit an acceptable resolution to the tension between cost effectiveness in test development and test-administration efficiency versus psychometric needs; (c) permit students to show growth in English language development from one year to the next; and (d) permit students to reach the ultimate target of *Full English Proficiency* within a realistic period of time.

The levels range from *Full English Proficiency (FEP)*, 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 may have some limited knowledge of English. The proficiency required for entry into the *Full English Proficiency* level is not synonymous with that of native-speaker proficiency in English; *FEP* students may function effectively and successfully in the school system while still exhibiting a non-native speaker accent, while making production errors (which typically would not impede communication), or while comprehending less than the full range of subtle meanings intended by a writer or speaker (again, with little negative effect on communication). By contrast, many aspects of an *Advanced* level proficiency and some aspects of an *Intermediate* and even of a *Beginners* level proficiency allow for the demonstration of an ability to function effectively in the school system, with less and less consistency and sophistication as one moves down the scale.

Focus of PLDs. AIR developed the PLDs according to guidelines established by the LEP-SCASS members at the project start-up Steering Committee meeting. The PLDs describe threshold points rather than the full range implied by a level; that is, each description characterizes what is minimally required for entry into a level. This is true of other second or foreign language scales of performance, including those of the American Council on the Teaching of Foreign Languages (the ACTFL Proficiency Guidelines), the Interagency Language Roundtable Proficiency Levels (ILR or Government Foreign Service scale), and the Council of Europe Proficiency Levels. The threshold approach is motivated by the constructs that are defined by the levels; that is, language skills that are cyclical and multidimensional and expanding patterns of learning rather than skills that are linear and one-dimensional and learning at a rate of complexity that is constant. For example, at level 1, a student may have no understanding of how to express present time in English; at level 2, a student may be able to express present time through the use of the present tense of some common verbs with simple adverbial present tense markers; and at level 3 and beyond, a student should have more extensive ways of expressing present time with the use of a greater range of verbs, with more sophisticated time markers, and with an ability to contrast present with other time references. The model is often described in the literature as an inverted pyramid in which progressively more language skill is required to attain entry into the next level. The PLD for the bottom level is an exception; it does not conform to the “threshold” requirement but rather is a description of a range, from zero knowledge and ability to just below what is minimally required for entry into level 2. The range implied by level 1, a *Pre-functional* or pre-communicative level (to continue the pyramid metaphor, the apex of the pyramid), is thus relatively straightforward.

PLD development procedures. An initial draft version of the PLDs for each of the four language skills was created at the project start-up Steering Committee meeting. Documents from the California State and New York State English Language Proficiency Levels, which represented substantive consideration of the issue of defining proficiency levels in the field of standards-aligned assessment for ESL, were consulted in this initial process. The draft version of the PLDs provided an important initial understanding and agreement of the type of characterization required at each of the five levels, reflecting a common understanding of the theoretical foundation for the descriptions. Particularly important was the determination of a “working definition” of level 5, *Fully English Proficient*. The initial draft version of the PLDs, however, lacked vertical and horizontal alignment. They were substantially reviewed and revised during the test development process to achieve alignment, both within domain and across the four domains. This review and revision process involved the following steps:

1. Analyze the original draft versions of the PLDs within a matrix to determine the performance indicators used to define each domain.
2. Assess the degree to which the indicators represented a complete and theoretically sound definition of the domain and aligned with test design specifications and scoring rubrics for constructed-response items, and then make appropriate revisions to the indicators.
3. Assess the degree to which the indicators were vertically aligned across the five levels and then make appropriate revisions.
4. Assess the degree to which the indicators were horizontally aligned across domains (particularly listening with reading, and speaking with writing) and then make appropriate revisions.
5. Analyze both listening and reading PLDs to determine what may be considered common for the creation of comprehension PLDs.
6. Review the content of all PLDs with internal AIR staff who are content experts but who are not related to the ELDA project.
7. Submit all PLDs to editorial review.

ELDA Operational Administration Procedures

Pre-Administration. At least one week prior to the ELDA testing window, all testing materials for all schools in a district are shipped to the District Assessment Coordinator (DAC). These shipments include the district packing list; the District Coordinator Manual; return shipping label(s); individual box(es) for each school; district overage materials; and a District Order Summary List (lists all materials packaged for each school). Upon receipt of these materials, the DAC is responsible for confirming that the shipment does indeed contain a District Coordinator Manual, return shipping label(s), and boxes for each school listed on the packing list. The DAC subsequently sends each school its respective box(es). Each school has a designated School Test Coordinator (STC), who is responsible for checking the contents of the school box(es), distributing assessment materials to test administrators, and maintaining the security of all documents.

ELDA K–2. The ELDA K–2 inventories are meant to be administered and scored by the classroom teacher, ESL teacher, or another qualified person familiar with the students and the skills being assessed. The STC supplies each test administrator with a set of ELDA K–2 Test Administrator Materials that includes a Test Administration Manual/Inventory, Teacher Support Materials, and enough 2-page scannable answer documents to record each student’s scores.

Before administering the inventories, teachers are advised to consult local curriculum authorities for guidance regarding the definition of terms related to passage length (short, medium, or long), level of support (low, moderate or high), and grade appropriateness of individual words and reading texts (common or high-frequency). ELDA is administered throughout the country, and because curriculum standards vary from state to state, it is impossible to establish a single set of universally applicable definitions.

The ELDA K–2 consists of four observation inventories, one each in Listening (7 rows), Reading (14 rows), Speaking (8 rows), and Writing (9 rows). Each row of a particular domain inventory reflects a discrete skill. Test administrators are advised to familiarize themselves with each of the inventories and the multiple behaviors they measure prior to administering the assessment.

In addition, it is necessary for test administrators to ensure that demographic and background information has been recorded for each student prior to beginning the assessment administration. In the event that a student does not have a precode label, page 1 of the answer document must be completed in full by the same person who is primarily responsible for most or all of the students' scoring. The Test Administration Manual provides step-by-step instructions for completing page 1 of the answer document.

ELDA 3–12. For grades 3–12, the STC delivers a set of ELDA 3–12 Test Administrator Materials to each test administrator. Test administrators must be trained to administer the ELDA and be in compliance with state and local assessment regulations and procedures. Each set of materials contains the following:

- Test Administration Manual
- Listening Prompt CD
- Speaking Prompt CD
- Speaking Scoring Guide
- Header Sheet
- Pre-printed student identification labels, if applicable; otherwise, generic answer document bar code labels
- Reading and Writing Test Booklets for students
- Listening and Speaking Test Booklets for students
- Student Answer Documents

Prior to the administration of the ELDA 3–12, STCs and/or test administrators apply a pre-printed student identification label to each answer folder. In the event that the student identification labels are not used, STCs and/or test administrators complete all required demographic information on each student's answer folder and apply a generic answer folder bar code label to the upper left-hand corner.

The room in which the assessment is administered should be such that each student has a work space large enough to accommodate an open test booklet and that students can be separated by a reasonable distance from one another during testing. Test administrators are advised to prepare the assessment room for the administration to ensure that (1) there is an area in the room for the secure storage of assessment materials when not in use, and (2) there are no visible resources (i.e., charts, maps, etc.) in the room that could inadvertently provide assistance to students.

Prior to administering the Listening and Speaking assessments, the test administrator ensures that a compact disc player is available to play the prompt recording. The player must provide adequate clarity and volume so that all students in the administration can hear the prompt recording clearly. Both the player and the prompt recording itself are tested prior to administration to ensure that the grade-level content of the recording is appropriate, that the volume is adequate, and that fresh batteries are inserted into the player if it is battery operated.

Administration, ELDA K–2. Each teacher administering and scoring the ELDA K and/or ELDA 1–2 is given a Test Administration Manual, a copy of the Teacher Support Materials, and scannable answer documents for each student being assessed. The Test Administration Manuals provide both an overview of administrative procedures and detailed instructions for the administration of each domain assessment.

All inventories and guidelines for completing them are included in the Test Administration Manual. The inventory scoring is based on student behavior and skill levels. Because typical behavior is not generally observed at a single point in time, teachers are advised to observe students over a period of two or more weeks before drawing a conclusion about a given student’s typical behavior. Consequently, more than one person often contributes to the administration and scoring process. When two or more people enter scores within an inventory, they are advised to discuss roles and responsibilities in advance.

Furthermore, the Test Administration Manual suggests that teachers observe each student in more than one setting (e.g., regular classroom, resource room, media center, playground, cafeteria, non-language instruction such as math or art, transitions, etc.) and in more than one type of interaction (e.g., with peers, familiar adults, unfamiliar adults, etc.). Because some inventory behaviors are less likely than others to occur naturally, teachers are directed to prompt such behaviors with the Teacher Support Materials or other more familiar resources.

The second page of the answer document contains bubbles for the test administrator to score a student on each inventory skill area (row) of the four inventories. As students at this grade-level range are emergent readers and writers, most student responses are rated by a test administrator who scores students based on their responses to specific prompts or behavior observations over the course of at least 2 weeks.

The behaviors and skill abilities being measured by the inventories represent varying degrees of mastery of content standards for each domain. Each behavior is scored on a 0–3 point scale. Descriptions of the score points 0–3 are given in each row. In general, a score point of 0 indicates that the student has not yet begun to acquire a skill or behavior. A score point of 1 represents minimal demonstration of the behavior set, evident either in a very limited range of behaviors or a wider range of behaviors at a very low level. A score point of 2 represents considerable progress in the behavior set, but not mastery of it, or mastery of some of the behaviors and skills in the set, but not all. Finally, a score point of 3 indicates mastery or near mastery of all or nearly all of the behaviors and skills listed in the set.

Each domain inventory contains rows that describe various behaviors and skill abilities. The rows on the inventories are numbered sequentially using the first letter of the domain followed by a two-digit number. For example, the Listening Inventory contains seven rows, labeled L01–L07. The content standard is listed first, followed by brief instructions for presentation and for assigning scores. The score point descriptors are then detailed for that standard.

Administration, ELDA 3–12. Each test administrator is given a set of ELDA 3–12 testing materials consisting of a Test Administration Manual; a Listening Prompt CD; a Speaking Prompt CD; a Speaking Scoring Guide; a Header Sheet; student identification labels; Reading

and Writing test booklets; Listening and Speaking test booklets; and Student Answer Documents. The Test Administration Manual provides both an overview of general administrative procedures and a set of detailed instructions for administering each assessment. To ensure standardized administration conditions, these instructions include scripted directions to be read aloud to the students. These oral scripts are to be reviewed by all test administrators prior to the assessment session. Test administrators are instructed to pay careful attention to which text is to be read aloud to the students (printed in bold type) and which text contains instructions to the administrator (printed in regular type).

To ensure that all students have sufficient time to complete the assessment, the ELDA is not timed, although suggested administration times for the short forms are provided in the Test Administration Manual for planning purposes and are reproduced in Table II.8, below. These are provided as a guideline for administrators, but they are suggestions only. If there are students who do not finish the test within the suggested time limit, you may have them complete the test in another room with a proctor.

Table II.8 Suggested ELDA Short Form Test Administration Times (in minutes)				
Cluster	Listening	Speaking	Reading	Writing
3–5	35	20	42	60
6–8	42	20	42	60
9–12	41	20	44	60

There is no recommended sequence for administering the Reading, Writing, and Listening tests. However, test administrators are advised to give the Speaking test last since they record the Speaking scores on the students’ answer documents.

Furthermore, it is the test administrator’s responsibility to provide breaks for students during the administration of the assessments. While the test administrator can use his or her own judgment as to how many breaks to provide and when to schedule them, the following guidelines apply:

- Before the test begins, students should be told when the break(s) will occur.
- The breaks should be timed with an available watch or clock.
- Students must remain in the assessment room during breaks; they should be encouraged to stand and stretch.
- No talking is permitted during breaks.
- No additional materials may be taken out during breaks.

At any time during the assessment administration, a student may leave the room for a restroom break, but to ensure assessment security, only one student should leave the room at any one time. The entire group may not use the restroom at the same time.

Reading assessment. The Reading section of the ELDA 3–12 tests how well students can read English. The test has three parts. Part 1 tests a student’s ability to understand information in short reading passages. Part 2 tests a student’s ability to understand written directions. Part 3 tests a student’s ability to understand information in longer reading passages.

Writing assessment. The Writing section of the ELDA 3–12 tests how well students can write in English. Part 1 requires that students write a response, in English, to a prompt. In Part 2, students read examples of student writing and then answer related multiple-choice questions. Similarly, in Part 3, students look at some examples of graphic organizers and then answer related multiple-choice questions.

Listening assessment. The Listening section of the ELDA 3–12 tests how well students listen and understand in English. Students listen to recorded stories and questions and respond to the questions by darkening circles in their answer documents. The Listening assessment for grades 3–5 consists of 4 parts; the Listening assessments for grades 6–8 and grades 9–12 consist of 5 parts. The recording is paused after each part of the assessment to make sure all students are ready to proceed.

On the recording, the narrator reads the entire content of the test booklet. In parts 1–4 of the Listening assessment, the stimulus material is read one time; in part 5, the stimulus material is read twice. The questions and answer choices are read only once in all parts. Students have 10 seconds to respond to each question after the narrator has read the last option.

Speaking assessment. The Speaking section of the ELDA 3–12 tests how well a student can speak in English. It is meant to be administered to students individually, rather than in a group setting. Students listen to recorded instructions and questions and are required to answer these questions in English within a prescribed period of time. Pictures in the Listening and Speaking test booklet, numbered to match the recorded questions, help students answer the questions.

The prompt recording has six sections. The first two sections provide opportunities for the student to practice. The assessment itself consists of English-Language Arts tasks; Mathematics, Science, and Technology tasks; and Social Studies tasks. The final section is a closing. Throughout the assessment, the test administrator scores the student’s responses in real time according to instructions provided in the Speaking Scoring Guide.

Post-Administration. After testing, the test administrators are responsible for returning the Test Administration Manual, Teacher Support Materials, and all answer documents (used and unused) to the designated STC, who returns all materials to the district in the same box in which they were originally delivered. Instructions for packing and taping the boxes securely are located in the Test Administration Manual. Ultimately, it is the responsibility of the DAC to account for all materials and, using the return shipping label(s) and the associated explicit instructions provided, return them to Measurement Incorporated (MI).

Security. Maintaining security is vital throughout the assessment process; security concerns are stressed throughout all communication with DACs, STCs, and test administrators. Persons involved with the assessment at any level are advised to follow district and state procedures for protecting secure assessment materials at all times.

- All ELDA materials—used and unused student answer folders, test booklets, prompt recordings, and speaking scoring guides—must be accounted for throughout the assessment and returned to MI after testing.

- Before and after an assessment administration, all assessment materials (both used and unused) must be kept in a secure location. Only those individuals authorized by district policy should have access to these materials. Under no circumstances should students have access to assessment materials before or after the assessment session.
- During each assessment administration, a test administrator must be in the room at all times. If students are allowed to leave the room while the assessment is in progress, they must first turn in their assessment materials. Administrators must account for all assessment materials before dismissing students.
- For the grades 3–12 ELDA, the ratio of test administrators to students is one administrator to 20 students in any assessment room for the Reading, Writing, and Listening tests, and one administrator to one student for the Speaking test.
- When assessment administration is complete and administrators have accounted for all materials, they should return them immediately to the STC.
- At any point, if the administrators or examiners believe that a violation of assessment security has occurred, they should follow the procedures established by the school district for handling alleged assessment security violations.

ELDA Scoring and Reporting Procedures

Scoring, ELDA K–2. All ELDA K–2 inventory items are scored in real time by the test administrator on a scale from 0–3 according to detailed instructions provided in the Test Administration Manual. The test administrator records the score on a machine-readable answer document, which is subsequently scanned at MI, and scores are assigned accordingly.

Scoring, ELDA 3–12. All ELDA 3–12 Reading and Listening items are multiple choice; students record one of four answer choices on machine-readable answer documents. These answer documents are scanned, and a score of 1 if the item is answered correctly or 0 if the item is answered incorrectly is assigned automatically with reference to approved answer keys.

ELDA 3–12 Writing Part 1 consists of a series of prompts to which students must respond in their own words and in their own writing. A trained reader evaluates each response and scores it according to specific scoring rubrics. Ten percent of the responses are independently evaluated by a second reader for purposes of scoring reliability. The final score assigned to the response is the score designated by the first reader. Writing Part 2 and Writing Part 3 items are multiple choice and are scored as described above. For the constructed-response portion of the Writing test, MI scoring center staff assigns scores on the basis of content as specified in the scoring rubrics. Blank responses (omissions) are scored as B, which result in a score of 0 for those items.

Speaking items are all constructed response. As described earlier in this document, students listen to pre-recorded instructions and questions and are required to answer these questions in English within a prescribed period of time. Pictures in the Listening and Speaking test booklet, numbered to match the recorded questions, help students answer the questions. Throughout the assessment, the test administrator scores the student’s responses in real time. The test administrator assigns a score ranging from 0–2 based on detailed instructions in the Speaking Scoring Guide. Alternately, the test administrator may assign a condition code to be used instead of—not in addition to—a score. These condition codes provide reasons for instances in which a

score was not assigned; these conditions include: inaudible/unclear; blank; refusal or excuse; and off topic/off task/completely in a language other than English.

For Speaking, the administrator or scorer records scores (0 to 2 or a condition code) on the answer document. If a condition code is entered, such coding indicates that the student was tested but did not receive credit. In such instances, the student receives a score of 0 for that item.

FOR OFFICIAL USE ONLY	
<input type="radio"/>	No longer enrolled
<input type="radio"/>	Emergency medical waiver
<input type="radio"/>	Student completed more than one answer folder
<input type="radio"/>	Student misclassified/no longer classified as ELL/LEP
<input type="radio"/>	Parent/Guardian refusal
<input type="radio"/>	Absent for entire/partial testing window

Answer document—For Official Use Only. The “For Official Use Only” field of the NE ELDA answer documents, reproduced above, encompasses six bubbles. Each bubble is associated with a potential reason for excluding the answer document from inclusion in analysis and reporting; the implications of darkened “For Official Use Only” bubbles are as follows:

- If the “No longer enrolled,” the “Emergency medical waiver,” the “Student completed more than one answer folder,” or the “Student misclassified/no longer classified as ELL/LEP” is darkened *and* the student answer document is completely blank, the student will receive a raw score and performance level of Not Tested (NT) for all four domains of ELDA.
- If the “Parent/Guardian Refusal” or the “Absent for entire/partial testing window” bubble is darkened *and* the student answer document is completely blank, the student will receive a raw score of 0 and a performance level of Refused to Test (RT) for all four domains of ELDA.
- If a student was issued a pre-ID label and no answer document was returned for that student, the student receives a raw score of 0 and a performance level of Refused to Test (RT) for all four domains of ELDA.
- If *any* bubble is darkened in the “For Official Use Only” field and some domains have scores recorded, the Nebraska Department of Education investigates to determine whether the student’s scores should be reported.
- If *no* bubble is darkened in the “For Official Use Only” field and no answers are marked for a given domain, the student receives a raw score of 0 and a performance level of 1 for that given domain.

Comprehension, Production, and Composite levels. For all grades, K–12, a student’s Comprehension score is based on his or her scores in both Reading and Listening, as shown in Table II.9, below.

Table II.9					
Rules for Combining Listening and Reading Levels to Yield a Comprehension Level					
ELDA K–2 and ELDA 3–12					
Reading	Listening				
	1	2	3	4	5
1	1	1	1	2	2
2	2	2	2	2	3
3	2	3	3	3	3
4	3	3	4	4	4
5	3	3	4	5	5

For ELDA K–2 only, a student’s Composite score is derived based on his or her Comprehension score from Table II.9 and his or her Production score from Table II.10.

Table II.10					
Rules for Combining Speaking and Writing Levels to Yield a Production Level					
ELDA K–2					
Writing	Speaking				
	1	2	3	4	5
1	1	1	1	2	2
2	2	2	2	2	3
3	2	3	3	3	3
4	3	3	4	4	4
5	3	3	4	5	5

The student’s Composite score, in grades K–2, is based on his or her scores in all four domains, as shown in Table II.11, below.

Table II.11					
Rules for Combining Comprehension and Production Levels to Yield a Composite Level					
ELDA K–2					
Production	Comprehension				
	1	2	3	4	5
1	1	1	2	2	3
2	1	2	2	3	3
3	2	2	3	3	4
4	2	3	3	4	4
5	3	3	4	4	5

For ELDA 3–12 only, a student’s Composite score is calculated according to a different set of procedures. Due to the development and administration of the new ELDA short forms, and according to recommendations of the ELDA Technical Advisory Committee (TAC), a compensatory scoring model was adopted. Under this model, a student’s Composite Performance Level is derived as follows:

Definitions:

- A student’s Component Performance Levels are the performance levels attained on each of the four ELDA components; these range from 1–5.
- A student’s Composite Performance Level reflects his or her performance on the ELDA as a whole.
- A student’s Intermediate Composite Performance Level (ICPL) is the weighted sum of a student’s Component Performance Levels; the ICPL ranges from 1–5.

Procedures:

1. Score the student’s ELDA test forms to obtain the four Component Performance Levels.
2. Multiply each Component Performance Level by the appropriate set of weights, according to Table II.12.
3. Add the weighted values; round the result down to the nearest hundredth.
4. Apply the appropriate set of ICPL cuts from Table II.13 to obtain the student’s Composite Performance Level.

Table II.12 ICPL Weights for Calculation of Composite Performance Level ELDA 3–12			
Domain	Grades 3–5	Grades 6–8	Grades 9–12
Speaking	0.03205	0.07435	0.04151
Listening	0.10466	0.09910	0.07175
Reading	0.43359	0.40512	0.46584
Writing	0.42970	0.42144	0.42090

Table II.13 ICPL Cut Scores for Calculation of Composite Performance Level ELDA 3–12			
Composite Performance Level	Grades 3–5	Grades 6–8	Grades 9–12
Level 1	NA	NA	NA
Level 2	1.75	1.69	1.70
Level 3	2.74	2.86	2.71
Level 4	3.71	3.84	3.74
Level 5	4.86	4.90	4.93

Reporting. Once all the items have been scored, MI’s Information Technology department generates four types of score reports to provide information at the student, school, and district levels: the Individual Student Report (ISR); the Student Roster; the District Summary Report; and the Demographic Report. All four reports provide information about performance in Listening, Speaking, Reading, and Writing. In addition, the reports include Comprehension and Composite proficiency levels, which are determined according to the rules outlined in Tables II.8 through II.13. Appendix D provides examples of the four types of reports, each of which is described briefly below.

- **Individual Student Report (ISR).** The ISR reports an individual student’s four scale scores and associated proficiency levels as well as Comprehension and Composite Performance Levels.
- **Student Roster.** The Student Roster summarizes the performance of all students tested in a class or school, reporting scale scores and proficiency levels.
- **District Summary Report.** The District Summary shows the numbers (N) and percentages (%) of students at each proficiency level; provides a mean (average) scale score for each test; and includes ranges of scale scores for each grade cluster and proficiency level.
- **Demographic Report.** The Demographic Report breaks down student performance by gender, ethnicity, time in ESL program, and first language. Groups are based on information students or teachers provided on the answer document. For each group, the report shows the total number of students taking all four tests; the percentage of students scoring at the Fully English Proficient level; and the mean scale score.

Quality control procedures for score reporting. MI thoroughly tests, reviews, and proofreads all score reports according to the following testing process (applicable to data files, electronic reports, and printed reports). MI uses structured testing methodologies to test all software programs that are used in the production of score reports. All scanning and scoring programs are fully tested and reviewed prior to the start of live material processing and are continually monitored throughout the process. MI staff develops and utilizes independent queries to validate all software programs and programmatically produced deliverables for reporting. Before reporting any results, each program is tested to ensure that:

- data are included or excluded as appropriate;
- programmatic calculations are performed accurately and according to the reporting rules;
- formatting rules are followed;
- reports are presented as designed with all appropriate fonts, footnotes, line separations, sections, and headings;
- valid values are verified;
- valid codes are included on all student records;
- correct scores are reflected and are attributed to the correct student;
- cluster scores are accurately aggregated and totaled; and
- appropriate student totals are reported in all aggregate reports.

III. Nebraska ELDA 2015 Operational Administration

Administration Summary

The Nebraska ELDA testing window runs annually from mid-February through March. In 2015, the shortened ELDA (Form 2) was administered to Nebraska ELLs between Monday, February 9, and Friday, March 20. Materials are provided to testing sites at least one week in advance of the testing window. Table III.1 illustrates the numbers of test booklets/answer documents scanned and scored in 2015.

Cluster	Number
1 (Kindergarten)	3,380
2 (Grades 1–2)	5,608
3 (Grades 3–5)	5,408
4 (Grades 6–8)	2,108
5 (Grades 9–12)	2,462
<i>Total</i>	<i>18,966</i>

Technical Characteristics

Data preparation and quality check. Prior to any reporting or analysis, all data is routinely run through a series of preparatory and quality assurance steps, as outlined below:

1. Answer keys are verified by IT and psychometric staff.
2. Answer documents are scanned and then go through data validation for the demographic data. Data validation involves executing a series of procedures, generally one for each field to be validated, to ascertain if the data entered meets criteria. A check on the litho code and barcode fields ensures that the document has not been previously scanned and is in the correct grade cluster group. Then individual field checks are done. For a field such as name, there is a basic check to see if the field is non-blank and does not have any double bubbles. For a field such as date of birth, checks are made to ensure the date is valid.
3. If an anomaly is detected, the field is flagged for Data Corrections. There, two operators will look at the document and enter any corrections. For example, if a field was not completely erased and another bubbled, the scanner may pick this up as a double bubble. A human can see where the bubble was erased and key in the correct value. If there is a difference between the results of the first and second operator, the document is flagged for a supervisor's correction. The programming staff, the project director, and the SDE may be contacted if there are questions that are not easily resolvable. An example of this might be when the student bubbles in a student ID but no name or date of birth.

4. Multiple-choice and constructed-response items are scored according to the procedures outlined earlier in this Technical Report.

At the conclusion of these steps, data are forwarded to the psychometrics department, according to a previously agreed-upon data layout scheme, for analysis. Because the psychometric analyses depend upon properly-prepared data, the analyses themselves provide an additional quality checkpoint; that is to say, the analyses conducted by the psychometricians will not make sense if the data are inaccurate.

Item and test analysis. Subsequent to data preparation, MI psychometricians conducted classical test theory analyses on all 2015 Nebraska operational data. For all grades and all domains, Appendices D through H illustrate 2015 operational item statistics, 2015 operational raw score frequency distributions, 2015 operational test summary statistics, 2015 operational raw score summary statistics, and 2015 proficiency level distribution respectively. Students whose answer documents indicated they were no longer enrolled, had an emergency medical waiver, completed more than one answer document, or were not classified as ELL/LEP were excluded from the analyses.

RT. Of the 18,508 student answer documents that were *not* excluded from analyses and reporting, 80 were assigned a performance level of RT for all four domains, indicating either (1) a parent/guardian refusal, (2) that a student was absent for some or all of the ELDA testing window, or (3) that a student was issued a pre-ID label but no answer document was returned. As described earlier, these students are assigned raw domain scores of 0. For the purposes of this technical report, these students are included in the item analysis.

Cluster	Number
1 (Kindergarten)	2
2 (Grades 1–2)	9
3 (Grades 3–5)	7
4 (Grades 6–8)	10
5 (Grades 9–12)	52
<i>Total</i>	<i>80</i>

Operational item statistics. Appendix D presents the results of classical item analysis for all grades and domains. Tables D.1 through D.8, which contain inventory data for clusters 1 (kindergarten) and 2 (grades 1–2), show item difficulty (mean score, with standard deviation), item discrimination (item-total correlation), number of respondents, and the distribution of scores at each score point for each item. The latter includes not only the percentage of respondents at each score point, from 0 to 3, but also the percentage of respondents who omitted (did not answer) an item; omits are counted as 0 for reporting purposes.

Tables D.9 through D.23 present the item statistics for clusters 3 (grades 3–5), 4 (grades 6–8) and 5 (grades 9–12), by domain. For Reading, Listening, and Writing (MC), on which the items are all multiple-choice, item difficulty is expressed as a p-value (the percentage of respondents

who answered the item correctly), and item discrimination is expressed as a point-biserial correlation. The response frequency distribution shows the percentage of respondents selecting each of the four answer choices; the percentage of respondents who omitted an item; and the percentage of respondents who “double-bubbled” (darkened more than one score point or answer choice; counted as 0 for reporting purposes).

For Writing (CR), data are presented in a format similar to that described for clusters 1 and 2 above—item difficulty (mean score, with standard deviation), item discrimination (item-total correlation), number of respondents, and score point distribution. The exception however is that, since *all* writing responses are scored (0–3 or 0–4), there is no separate column for omitted items. (Double-bubbles, of course, are not relevant for constructed-response items.)

The Speaking data are presented in a slightly different form. Because the Speaking responses are scored on a 0–2 scale, the tables display item difficulty (mean score, with standard deviation), item discrimination (item-total correlation), number of respondents, and score-point distributions for each item. There is also a separate column for omitted items. Should a student neglect to respond to a Speaking item, the administrator is to indicate a condition code to explain *why* or *in what way* a student did not answer. A condition code is assigned *instead* of—not in addition to—a score. The condition codes include:

- A –inaudible/unclear
- B –blank
- C –refusal or excuse (“I don’t want to answer;” “I don’t know”)
- D –off topic/off task/completely in language other than English

The distribution of condition codes (which, again, pertain *only* to omitted items) is displayed just below the Speaking item statistics for a given cluster. While the ELDA Speaking Scoring Guide for grades 3–12 implies that the test administrator is to indicate a condition code if a speaking item is omitted, no such requirement is made clear; consequently, the degree to which condition codes were assigned to explain omitted items varied from item to item and cluster to cluster. Of the condition codes that *were* assigned, the majority across all clusters were coded B (blank) or D (off topic/off task/completely in language other than English).

Operational raw score frequency distributions. Appendix E presents the raw score frequencies and cumulative frequencies for all grades and domains.

Operational test summary statistics. Appendix F provides reliability coefficients and test statistics for all ELDA assessments administered in 2015. For each cluster and domain test, the table illustrates the number of respondents, the maximum possible raw score, the mean raw score and associated standard deviation, and the reliability coefficient and associated standard error of measurement. Using SAS version 9.1 (SAS Institute, Inc., 2003), MI estimated reliability coefficients with Cronbach’s alpha. Reliability coefficients for all domains in grades K–2 range from 0.93 to 0.96. ELDA 3–12 reliability coefficients were also strong, ranging from a low of 0.80 (Cluster 3 Writing) to a high of 0.96 (Clusters 4 and 5 Speaking). Reliability coefficients for the ELDA Writing assessment have historically been somewhat lower—yet still healthy and strong—than those for the other domains.

Operational raw score summary statistics. Appendix G presents raw score summary statistics by gender and ethnicity for each grade cluster and test domain.

Performance levels. As discussed in detail earlier in this document, ELDA distinguishes five achievement levels: Pre-functional (1), Beginning (2), Intermediate (3), Advanced (4), and Fully English Proficient (5). In addition, as indicated earlier, a small percentage of examinees were assigned performance levels of RT. Tables H.1 through H.5, attached as Appendix H, illustrate the percentage of students achieving a given level of proficiency, by cluster and domain.

IV. References

- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: AERA.
- American Institutes for Research (2005, October). *ELDA English Language Proficiency Standards and Test and Item Specifications for Grades 3–12*. Washington, D.C.: Author.
- American Institutes for Research and Measurement Incorporated (2005, October). *ELDA Technical Report: 2004 Field Test Administration*. Washington, D.C.: American Institutes for Research.
- Bunch, M. B. (2006, February). *Final Report on ELDA Standard Setting*. Durham, N.C.: Measurement Incorporated.
- Bunch, M. B. (2006, February). *ELDA K–2 Development*. Durham, N.C.: Measurement Incorporated.
- Bunch, M. B., & Joldersma, K. (2006). *Supplement to the ELDA K–2 Technical Report*. Durham, N.C.: Measurement Incorporated.
- Center for Advanced Studies in Measurement and Assessment, University of Iowa. *LEGS* [Computer Software]. Iowa City, IA: Author.
- Cook, H. G. (2005, December). *Aligning English Language Proficiency Tests to English Language Learning Standards*. Madison, WI: Wisconsin Center for Education Research.
- Council of Chief State School Officers (2006, Spring). *ELDA K–2 Technical Manual: Spring 2006*. Washington, D.C.: Author.
- Haladyna, T. M. (2009, September). *A Technical Report: Four Studies Intended to Gather Information for Reducing the Administration Time of ELDA*. Phoenix, AZ: Author.
- Jaeger, R. M. (1989). *Certification of Student Competence*. In R. L. Linn (Ed.), *Educational Measurement* (3rd Edition), (pp. 485–514). New York: American Council on Education & Macmillan Publishing Co.
- Linacre, J. M. (2008). *Winsteps* (Version 3.65.0) [Computer Software]. Chicago: Winsteps.com.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Measurement Incorporated (2009, January). *Tennessee English Language Development Assessment Placement Test: Technical Report, 2008 Field Test and Operational Development*. Durham, N.C.: Author.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), in *Educational Measurement* (3rd. edition, page 13–103). New York: American Council on Education and Macmillan.
- SAS Institute, Inc. (2003). SAS 9.1. Cary, N.C.: SAS Institute, Inc.
- Wolf, M. K. Kao, J., Griffin, N., Herman, J.L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008, January). *Issues in assessing English Language Learners: English Language Proficiency measures and accommodation uses practice review (part 2 of 3)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Appendix A

2015 Conversion Tables

Table A.1
Raw-Theta-Scale Score Conversions
Cluster 1 Kindergarten Reading

RS	Theta	Scale	Level
0	-9.42	38	1
1	-7.96	58	1
2	-6.99	71	1
3	-6.41	79	1
4	-5.98	85	1
5	-5.65	90	1
6	-5.37	93	1
7	-5.12	97	1
8	-4.89	100	2
9	-4.68	103	2
10	-4.48	106	2
11	-4.30	108	2
12	-4.12	111	2
13	-3.95	113	2
14	-3.78	115	2
15	-3.62	117	2
16	-3.46	120	2
17	-3.30	122	2
18	-3.15	124	2
19	-2.99	126	2
20	-2.84	128	3
21	-2.69	130	3
22	-2.54	132	3
23	-2.40	134	3
24	-2.25	136	3
25	-2.10	138	3
26	-1.95	140	3
27	-1.79	142	3
28	-1.64	144	3
29	-1.48	147	3
30	-1.32	149	3
31	-1.15	151	3
32	-0.97	154	3
33	-0.78	156	3
34	-0.58	159	3
35	-0.36	162	3
36	-0.12	165	4
37	0.15	169	4
38	0.46	173	4
39	0.85	178	4
40	1.36	185	5
41	2.17	196	5
42	3.46	214	5

Table A.2
Raw-Theta-Scale Score Conversions
Cluster 1 Kindergarten Writing

RS	Theta	Scale	Level
0	-10.11	38	1
1	-8.70	54	1
2	-7.68	66	1
3	-6.91	74	1
4	-6.23	82	1
5	-5.63	89	1
6	-5.10	95	1
7	-4.63	100	2
8	-4.20	105	2
9	-3.81	109	2
10	-3.44	113	2
11	-3.09	117	2
12	-2.76	121	2
13	-2.43	125	2
14	-2.10	128	2
15	-1.77	132	2
16	-1.44	136	3
17	-1.09	140	3
18	-0.73	144	3
19	-0.36	148	3
20	0.03	152	3
21	0.44	157	4
22	0.87	162	4
23	1.34	167	4
24	1.88	173	4
25	2.59	181	4
26	3.69	193	5
27	5.25	211	5

Table A.3
Raw-Theta-Scale Score Conversions
Cluster 1 Kindergarten Listening

RS	Theta	Scale	Level
0	-11.60	60	1
1	-10.14	73	1
2	-9.02	84	1
3	-8.11	92	1
4	-7.26	100	2
5	-6.49	107	2
6	-5.81	114	2
7	-5.18	119	2
8	-4.57	125	2
9	-3.98	131	3
10	-3.38	136	3
11	-2.79	142	3
12	-2.19	147	3
13	-1.59	153	3
14	-0.98	158	3
15	-0.34	164	3
16	0.34	171	4
17	1.06	177	4
18	1.82	184	4
19	2.64	192	5
20	3.69	202	5
21	5.10	215	5

Table A.4
Raw-Theta-Scale Score Conversions
Cluster 1 Kindergarten Speaking

RS	Theta	Scale	Level
0	-9.55	49	1
1	-8.19	63	1
2	-7.23	72	1
3	-6.50	80	1
4	-5.82	87	1
5	-5.15	93	1
6	-4.49	100	2
7	-3.88	106	2
8	-3.33	112	2
9	-2.81	117	2
10	-2.33	122	2
11	-1.85	127	2
12	-1.37	131	3
13	-0.87	136	3
14	-0.34	142	3
15	0.22	148	3
16	0.83	154	3
17	1.47	160	3
18	2.13	167	4
19	2.79	173	4
20	3.48	180	4
21	4.24	188	4
22	5.10	197	5
23	6.19	208	5
24	7.64	222	5

Table A.5
Raw-Theta-Scale Score Conversions
Cluster 2 Grades 1–2 Reading

RS	Theta	Scale	Level
0	-8.83	46	1
1	-7.58	63	1
2	-6.80	74	1
3	-6.31	81	1
4	-5.93	86	1
5	-5.61	90	1
6	-5.32	94	1
7	-5.05	98	1
8	-4.80	101	1
9	-4.56	105	1
10	-4.32	108	2
11	-4.09	111	2
12	-3.86	114	2
13	-3.64	117	2
14	-3.42	120	2
15	-3.20	123	2
16	-2.99	126	2
17	-2.79	129	2
18	-2.58	132	2
19	-2.38	134	2
20	-2.18	137	2
21	-1.99	140	2
22	-1.79	142	3
23	-1.60	145	3
24	-1.40	148	3
25	-1.21	150	3
26	-1.01	153	3
27	-0.80	156	3
28	-0.60	159	3
29	-0.38	162	3
30	-0.16	165	3
31	0.07	168	4
32	0.31	171	4
33	0.55	174	4
34	0.81	178	4
35	1.08	182	4
36	1.37	186	4
37	1.68	190	4
38	2.03	195	4
39	2.43	200	5
40	2.95	207	5
41	3.76	218	5
42	5.03	236	5

Table A.6
Raw-Theta-Scale Score Conversions
Cluster 2 Grades 1–2 Writing

RS	Theta	Scale	Level
0	-10.68	32	1
1	-9.23	48	1
2	-8.20	60	1
3	-7.51	68	1
4	-6.94	74	1
5	-6.45	80	1
6	-5.98	85	1
7	-5.51	90	1
8	-5.04	95	2
9	-4.57	101	2
10	-4.10	106	2
11	-3.65	111	2
12	-3.23	116	2
13	-2.82	120	2
14	-2.43	125	2
15	-2.04	129	2
16	-1.63	134	2
17	-1.20	138	3
18	-0.75	144	3
19	-0.27	149	3
20	0.23	155	3
21	0.74	160	4
22	1.26	166	4
23	1.81	172	4
24	2.44	179	4
25	3.21	188	4
26	4.27	200	5
27	5.71	216	5

Table A.7
Raw-Theta-Scale Score Conversions
Cluster 2 Grades 1–2 Listening

RS	Theta	Scale	Level
0	-10.86	67	1
1	-9.48	79	1
2	-8.51	88	1
3	-7.78	95	1
4	-7.11	101	1
5	-6.44	108	1
6	-5.69	115	2
7	-4.87	122	2
8	-4.08	130	2
9	-3.40	136	2
10	-2.82	141	2
11	-2.27	146	3
12	-1.70	152	3
13	-1.08	158	3
14	-0.36	164	3
15	0.42	171	3
16	1.21	179	4
17	1.97	186	4
18	2.71	193	4
19	3.49	200	5
20	4.49	209	5
21	5.88	222	5

Table A.8
Raw-Theta-Scale Score Conversions
Cluster 2 Grades 1–2 Speaking

RS	Theta	Scale	Level
0	-8.84	56	1
1	-7.47	70	1
2	-6.51	80	1
3	-5.82	87	1
4	-5.23	93	1
5	-4.69	98	1
6	-4.17	103	1
7	-3.69	108	1
8	-3.21	113	2
9	-2.76	117	2
10	-2.32	122	2
11	-1.89	126	2
12	-1.44	131	2
13	-0.96	136	3
14	-0.41	141	3
15	0.24	148	3
16	1.01	155	3
17	1.82	164	3
18	2.55	171	4
19	3.22	178	4
20	3.89	184	4
21	4.61	192	4
22	5.43	200	5
23	6.54	211	5
24	8.00	226	5

Table A.9
Raw-Theta-Scale Score Conversions
Cluster 3 Grades 3–5 Reading
(short form 2)

RS	Theta	Scale	Level
0	-5.85	93	1
1	-4.62	122	1
2	-3.81	142	1
3	-3.30	154	1
4	-2.92	170	1
5	-2.61	207	1
6	-2.35	239	1
7	-2.11	268	1
8	-1.91	293	1
9	-1.72	316	1
10	-1.54	338	1
11	-1.38	357	1
12	-1.22	377	1
13	-1.07	395	1
14	-0.92	413	1
15	-0.78	430	1
16	-0.64	448	1
17	-0.51	463	2
18	-0.37	480	2
19	-0.24	496	2
20	-0.10	513	2
21	0.03	529	2
22	0.17	546	2
23	0.32	565	2
24	0.46	582	3
25	0.62	601	3
26	0.78	621	3
27	0.95	641	3
28	1.13	663	4
29	1.34	689	4
30	1.57	717	4
31	1.84	750	4
32	2.17	790	5
33	2.62	845	5
34	3.36	902	5
35	4.64	931	5

Table A.10
Raw-Theta-Scale Score Conversions
Cluster 3 Grades 3–5 Listening
(short form 2)

RS	Theta	Scale	Level
0	-5.55	96	1
1	-4.18	138	1
2	-3.43	160	1
3	-2.97	195	1
4	-2.63	236	1
5	-2.35	269	1
6	-2.11	298	1
7	-1.90	323	1
8	-1.71	345	1
9	-1.53	367	1
10	-1.36	387	1
11	-1.20	406	1
12	-1.05	424	1
13	-0.90	442	1
14	-0.75	460	2
15	-0.61	476	2
16	-0.47	493	2
17	-0.33	510	2
18	-0.19	526	2
19	-0.05	543	2
20	0.09	560	3
21	0.23	576	3
22	0.38	594	3
23	0.53	612	3
24	0.68	630	3
25	0.84	649	4
26	1.01	669	4
27	1.20	692	4
28	1.39	714	4
29	1.61	740	5
30	1.85	769	5
31	2.14	804	5
32	2.49	845	5
33	2.95	900	5
34	3.70	918	5
35	4.60	935	5

Table A.11
Raw-Theta-Scale Score Conversions
Cluster 3 Grades 3–5 Writing
(short form 2)

RS	Theta	Scale	Level
0	-4.52	99	1
1	-3.79	175	1
2	-3.06	250	1
3	-2.62	296	1
4	-2.30	330	1
5	-2.03	357	1
6	-1.81	381	1
7	-1.61	401	1
8	-1.42	421	1
9	-1.25	439	1
10	-1.08	456	2
11	-0.92	473	2
12	-0.76	490	2
13	-0.60	506	2
14	-0.44	523	2
15	-0.27	540	2
16	-0.11	557	2
17	0.07	576	2
18	0.25	594	3
19	0.44	614	3
20	0.65	635	3
21	0.87	659	3
22	1.11	684	4
23	1.39	713	4
24	1.73	748	4
25	2.15	792	4
26	2.74	853	4
27	3.72	908	5
28	4.69	933	5

Table A.12
Raw-Theta-Scale Score Conversions
Cluster 3 Grades 3–5 Speaking
(short form 2)

RS	Theta	Scale	Level
0	-4.51	163	1
1	-3.29	275	1
2	-2.57	341	1
3	-2.14	381	1
4	-1.81	411	1
5	-1.55	435	1
6	-1.33	456	2
7	-1.12	475	2
8	-0.93	492	2
9	-0.76	508	2
10	-0.58	524	2
11	-0.42	539	2
12	-0.25	555	3
13	-0.07	571	3
14	0.10	587	3
15	0.29	604	3
16	0.49	623	3
17	0.71	643	3
18	0.95	665	3
19	1.23	691	4
20	1.55	720	4
21	1.96	758	4
22	2.51	809	5
23	3.39	890	5
24	4.74	940	5

Table A.13
Raw-Theta-Scale Score Conversions
Cluster 4 Grades 6 – 8 Reading
(short form 2)

RS	Theta	Scale	Level
0	-5.04	112	1
1	-3.92	139	1
2	-3.17	157	1
3	-2.72	194	1
4	-2.38	235	1
5	-2.11	268	1
6	-1.87	298	1
7	-1.67	322	1
8	-1.48	345	1
9	-1.31	366	1
10	-1.15	385	1
11	-0.99	405	1
12	-0.85	422	1
13	-0.71	439	1
14	-0.57	456	1
15	-0.44	472	2
16	-0.31	488	2
17	-0.18	504	2
18	-0.05	520	2
19	0.08	535	2
20	0.21	551	2
21	0.34	567	2
22	0.48	584	2
23	0.62	601	2
24	0.76	618	3
25	0.91	637	3
26	1.07	656	3
27	1.24	677	3
28	1.43	700	4
29	1.63	724	4
30	1.87	754	4
31	2.14	787	4
32	2.47	827	4
33	2.93	883	5
34	3.67	909	5
35	4.80	935	5

Table A.14
Raw-Theta-Scale Score Conversions
Cluster 4 Grades 6 – 8 Listening
(short form 2)

RS	Theta	Scale	Level
0	-4.86	117	1
1	-3.65	154	1
2	-2.90	204	1
3	-2.45	257	1
4	-2.11	298	1
5	-1.84	330	1
6	-1.61	357	1
7	-1.40	382	1
8	-1.22	404	1
9	-1.05	424	1
10	-0.89	443	1
11	-0.73	462	1
12	-0.59	479	1
13	-0.45	495	1
14	-0.31	512	1
15	-0.18	527	1
16	-0.05	543	1
17	0.08	558	2
18	0.21	574	2
19	0.34	589	2
20	0.47	605	2
21	0.60	620	2
22	0.74	637	3
23	0.88	654	3
24	1.03	671	3
25	1.18	689	3
26	1.34	708	3
27	1.51	729	4
28	1.70	751	4
29	1.91	776	4
30	2.14	804	4
31	2.42	837	5
32	2.75	876	5
33	3.21	910	5
34	3.95	923	5
35	4.93	940	5

Table A.15
Raw-Theta-Scale Score Conversions
Cluster 4 Grades 6–8 Writing
(short form 2)

RS	Theta	Scale	Level
0	-3.91	163	1
1	-3.20	236	1
2	-2.49	310	1
3	-2.06	354	1
4	-1.75	386	1
5	-1.50	412	1
6	-1.29	435	1
7	-1.10	454	1
8	-0.92	473	1
9	-0.76	490	1
10	-0.60	506	1
11	-0.44	522	1
12	-0.29	538	1
13	-0.14	554	2
14	0.01	570	2
15	0.17	586	2
16	0.33	602	2
17	0.49	620	2
18	0.66	637	2
19	0.85	657	3
20	1.05	677	3
21	1.26	699	3
22	1.50	724	4
23	1.77	753	4
24	2.09	786	4
25	2.49	827	4
26	3.03	883	4
27	3.88	912	5
28	4.73	934	5

Table A.16
Raw-Theta-Scale Score Conversions
Cluster 4 Grades 6–8 Speaking
(short form 2)

RS	Theta	Scale	Level
0	-4.38	175	1
1	-3.46	260	1
2	-2.67	332	1
3	-2.18	377	1
4	-1.81	411	1
5	-1.51	439	1
6	-1.24	464	2
7	-1.01	485	2
8	-0.79	505	2
9	-0.59	524	2
10	-0.40	541	2
11	-0.22	558	2
12	-0.04	574	2
13	0.15	592	2
14	0.33	608	2
15	0.52	626	3
16	0.72	644	3
17	0.94	664	3
18	1.17	685	3
19	1.44	710	3
20	1.76	740	4
21	2.15	776	4
22	2.68	824	4
23	3.51	900	5
24	4.55	934	5

Table A.17
Raw-Theta-Scale Score Conversions
Cluster 5 Grades 9–12 Reading
(short form 2)

RS	Theta	Scale	Level
0	-4.58	123	1
1	-3.18	157	1
2	-2.44	228	1
3	-2.00	282	1
4	-1.67	322	1
5	-1.41	354	1
6	-1.18	382	1
7	-0.98	406	1
8	-0.80	428	1
9	-0.64	448	1
10	-0.49	466	1
11	-0.34	484	1
12	-0.20	501	1
13	-0.07	517	1
14	0.06	533	1
15	0.19	549	2
16	0.31	563	2
17	0.43	578	2
18	0.55	593	2
19	0.68	609	2
20	0.80	623	2
21	0.93	639	3
22	1.06	655	3
23	1.19	671	3
24	1.33	688	3
25	1.47	705	3
26	1.62	723	4
27	1.79	744	4
28	1.96	765	4
29	2.16	789	4
30	2.38	816	4
31	2.65	849	4
32	2.97	888	5
33	3.42	903	5
34	4.15	920	5
35	5.44	949	5

Table A.18
Raw-Theta-Scale Score Conversions
Cluster 5 Grades 9–12 Listening
(short form 2)

RS	Theta	Scale	Level
0	-4.81	119	1
1	-3.37	148	1
2	-2.62	237	1
3	-2.17	290	1
4	-1.83	331	1
5	-1.56	363	1
6	-1.33	390	1
7	-1.12	415	1
8	-0.93	438	1
9	-0.76	458	1
10	-0.60	477	1
11	-0.45	495	1
12	-0.31	512	1
13	-0.17	529	1
14	-0.03	545	1
15	0.10	561	2
16	0.23	576	2
17	0.36	592	2
18	0.49	607	2
19	0.62	623	2
20	0.75	638	3
21	0.89	655	3
22	1.02	670	3
23	1.16	687	3
24	1.31	705	3
25	1.46	723	3
26	1.62	742	4
27	1.80	763	4
28	1.99	786	4
29	2.19	810	4
30	2.43	838	4
31	2.70	870	5
32	3.04	907	5
33	3.50	915	5
34	4.24	928	5
35	5.48	950	5

Table A.19
Raw-Theta-Scale Score Conversions
Cluster 5 Grades 9–12 Writing
(short form 2)

RS	Theta	Scale	Level
0	-4.57	93	1
1	-3.81	173	1
2	-3.04	253	1
3	-2.56	302	1
4	-2.20	340	1
5	-1.91	370	1
6	-1.66	396	1
7	-1.43	419	1
8	-1.23	441	1
9	-1.04	460	1
10	-0.86	479	1
11	-0.69	496	1
12	-0.53	513	2
13	-0.37	530	2
14	-0.21	547	2
15	-0.05	563	2
16	0.11	580	2
17	0.27	596	2
18	0.43	613	2
19	0.60	631	3
20	0.77	649	3
21	0.95	668	3
22	1.15	688	3
23	1.35	709	3
24	1.58	732	4
25	1.83	758	4
26	2.11	788	4
27	2.45	823	4
28	2.88	868	5
29	3.47	902	5
30	4.44	927	5
31	5.41	951	5

Table A.20
Raw-Theta-Scale Score Conversions
Cluster 5 Grades 9–12 Speaking
(short form 2)

RS	Theta	Scale	Level
0	-3.79	229	1
1	-2.92	309	1
2	-2.18	377	1
3	-1.73	419	1
4	-1.39	450	1
5	-1.11	476	1
6	-0.87	498	1
7	-0.65	518	1
8	-0.45	536	1
9	-0.26	554	1
10	-0.07	571	2
11	0.12	589	2
12	0.31	606	2
13	0.50	624	2
14	0.70	642	2
15	0.91	661	3
16	1.14	683	3
17	1.38	705	3
18	1.63	728	3
19	1.91	753	3
20	2.22	782	4
21	2.58	815	4
22	3.05	858	5
23	3.80	910	5
24	4.79	941	5

Figure A.1 provides the algorithms for the theta-to-scale score conversions. The trim algorithms for grades 3–12 (shortened forms) are the means by which values at the extreme upper and lower ends of the scale are trimmed back to prevent negative scale scores or scores above 1,000.

Algorithms	
<u>Listening</u>	
K–2	$SS = 167.543 + 9.3 * \text{Theta}$
3–12	$SS = 450 + 119.0476 * (\text{Theta} + 0.83)$
	<i>Lower Trim (Theta < -3.21)</i> $SS = 100 + 30.29 * (\text{Theta} + 5.42)$
	<i>Upper Trim (Theta > 3.00)</i> $SS = 906 + 17.81 * (\text{Theta} - 3.00)$
<u>Reading</u>	
K–2	$SS = 166.802 + 13.661 * \text{Theta}$
3–12	$SS = 450 + 121.9512 * (\text{Theta} + 0.62)$
	<i>Lower Trim (Theta < -2.987)</i> $SS = 100 + 23.81152 * (\text{Theta} + 5.5627)$
	<i>Upper Trim (Theta > 3.021)</i> $SS = 894 + 22.96023 * (\text{Theta} - 3.021)$
<u>Speaking</u>	
K–2	$SS = 145.264 + 10.081 * \text{Theta}$
3–12	$SS = 450 + 91.95402 * (\text{Theta} + 1.39)$
	<i>No Lower Trim</i>
	<i>Upper Trim (Theta > 3.5)</i> $SS = 900 + 32.051 * (\text{Theta} - 3.5)$
<u>Writing</u>	
K–2	$SS = 152.023 + 11.236 * \text{Theta}$
3–12	$SS = 450 + 103.8961 * (\text{Theta} + 1.14)$
	<i>No Lower Trim</i>
	<i>Upper Trim (Theta > 3.12):</i> $SS = 893 + 25.333 * (\text{Theta} - 3.12)$

Figure A.1 Scale score conversion algorithms.

Appendix B

Performance Level Descriptors

**Table B.1
Indicators Defining Each ELDA Language Domain**

Performance Indicators	Receptive Skills			Productive Skills	
	Listening	Reading	Comprehension	Speaking	Writing
Text types • content area/ non-content area		✓			✓
Discourse types • content area/ non-content area	✓		✓	✓	
Speech types • connect, tell, expand, reason • content area/ non-content area				✓	
Forming a general understanding • main idea, theme, problem, conflict, plot, character, event, mood, message, purpose	✓	✓	✓		
Developing an understanding • details	✓	✓	✓		
Linking information • communicator point of view • inference, conclusion, evaluation	✓	✓	✓		
Vocabulary and structure • academic/school-social • formal/informal	✓	✓	✓		
Use of vocabulary • academic/school-social • formal/informal				✓	✓
Text-level structure • organization • logic of argument • cohesive devices				✓	✓
Sentence-level structure • tense • modality • word order • inflection				✓	✓
Mechanics • punctuation • spelling • capitalization					✓
Fluency • creativity, spontaneity, flexibility • pronunciation				✓	

Performance Level Definitions – Grades K–2

READING

Level 5 (Fully English Proficient). Students at this level participate in reading activities with little teacher support at a level comparable to their English-speaking peers. They read for different purposes across a variety of text types. They have an increasing range of receptive nonacademic and academic vocabulary that allows them to read with greater fluency. They understand multiple word meanings. They have greater comprehension as a result of their increasing control of the structures of English. They can make connections between what they read and other experiences and tasks.

Level 4 (Advanced). Students at this level can read familiar text with little teacher or visual support. However, they still need those supports when reading to comprehend unfamiliar text. They can apply their phonemic awareness skills to read more complicated text. They have oral fluency and use self-monitoring and self-correction strategies when necessary. They use pre-, during-, and post- reading strategies but still need teacher prompting to use these skills. They can identify all story elements and can recognize cause and effect relationships in the texts they read. They make connections between the texts they read and themselves, the world, and other texts. They comprehend text in read aloud and can participate in the majority of read-aloud activities. They are beginning to read across text types and apply what they read to other activities.

Level 3 (Intermediate). Students at this level are developing phonemic awareness skills that allow them to read single words and simple text with comprehension. Reading is aided by visual and teacher supports. At this stage, oral reading is hesitant and difficult to understand due to a lack of oral language proficiency. These students have a small repertoire of high frequency words. They are beginning to use simple reading strategies and to make self, world, and other text connections to the text they are reading. They comprehend simple sentence structure and sentences with simple compounding. They recognize that words serve different functions, have multiple meanings, and have both synonyms and antonyms. In read aloud, with teacher support, they can identify some story elements and retell the majority of the story.

Level 2 (Beginning). Students at this level begin to identify the names of both upper and lower case letters of the alphabet. They use juncture to identify where words begin and end. They begin to recognize that words serve different functions (e.g., nouns, verbs). They can follow multistep directions depicted graphically. During read aloud, they get meaning primarily from pictures and the teacher’s tone of voice and gestures.

Level 1 (Pre-functional). Students at this level demonstrate an understanding of concepts of print (e.g., front-to-back, top-to-bottom, left-to-right) and begin to track print. They can distinguish letters from other symbolic representations. They can follow one-step directions depicted graphically. They can imitate the act of reading (e.g., holding a book and turning pages); however, they get meaning only through pictures.

Performance Level Definitions – Grades K–2 (cont’d)

LISTENING

Level 5 (Fully English Proficient). Students at this level understand most grade-level appropriate content-area and school/social speech. They understand the main ideas and relevant details of extended discussions or oral presentations on a range of familiar and unfamiliar topics comparable to a native English speaker at the same grade level. They are capable of making interpretations of what they hear. They understand most of the complex structures of spoken English relative to their grade level. They have a broad range of vocabulary, including idiomatic language, relating to both content areas and school/social environments.

Level 4 (Advanced). Students at this level understand conversations in most school/social settings. They understand main ideas and significant relevant details of extended discussions or presentations on familiar and relevant academic topics. They are able to comprehend conversations and orally-delivered texts involving description and narration in different time frames or conditions. They understand most of the basic language forms of spoken English, including timeless conditionals and sentences using clauses and phrases. They are able to understand cohesive devices to follow the sequence in an oral presentation or text. They comprehend most grade-level vocabulary and idioms, especially school/social environments, and are beginning to develop a wide range of academic vocabulary related to content areas, with limited supports such as visuals and rephrasing. They understand multiple meanings of words and can use context clues to understand messages.

Level 3 (Intermediate). Students at this level understand sentence-length statements and questions that include recombinations of learned language structures and on a variety of social and academic topics. They understand simple and compound sentences. They understand time through the use of simple tenses that may not be supported by adverbials of time. They are able to understand multistep directions. They also understand the difference between statements and questions by intonation, word order, and interrogative words. They understand and are able to identify main ideas and some details from conversations and simple/age appropriate orally-delivered text, usually with visual supports in familiar communicative situations and in academic content areas. They begin to interpret meaning from conversations and orally-delivered text, making predictions and drawing conclusions. They understand some idioms, mostly related to school/social environments, and have key vocabulary from content areas. They are aware of cohesive devices but may not be able to use them to follow the sequence of thought in an oral text.

Level 2 (Beginning). Students at this level understand simple, short statements and questions on a well-known topic within a familiar context. Tense is understood through the use of adverbials or situation rather than inflectional endings. They are able to follow simple multistep directions. They identify the main idea and some details of short conversations or simple orally-delivered text on a familiar topic. They understand basic grammatical structures and vocabulary in the school and social environment. Students at this level still need frequent repetition and rephrasing. They understand what they have heard but not variations or recombinations of what they have heard.

Level 1 (Pre-functional). Students at this level may understand some isolated words (particularly school and social environment vocabulary), some high frequency social conventions, and simple directions (single word or short phrase), commands and questions. They rely on non-verbal cues such as gestures and facial expressions and require frequent repetition and rephrasing to understand spoken language. They need strong situational support to understand most oral language.

Performance Level Definitions – Grades K–2 (cont’d)

SPEAKING

Level 5 (Fully English Proficient). Students who are ready to enter Level 5, Fully English Proficient, can supply coherent, unified and appropriately sequenced responses to an interlocutor. They use a variety of devices to connect ideas logically. They understand and can use a range of complex and simple grammatical structures, as appropriate for topic and type of discourse. Their grammar and vocabulary is comparable to that of a minimally proficient native English speaker—grammar errors very seldom impede communication and their range of school-social and academic vocabulary allows a precision of speech comparable to a native English speaker. They can effectively engage in non-interactive speech. They can use language effectively to connect, tell, expand, and reason. They show flexibility, creativity, and spontaneity in speech in a variety of contexts.

Level 4 (Advanced). Students entering proficiency Level 4, the Advanced level, are able to restructure the language they know to meet the creative demands of most social and academic situations. They can supply mostly coherent, unified and appropriately sequenced responses to an interlocutor. They use some devices to connect ideas logically and they use a range of grammatical structures. They make some errors in modality, tense, agreement, pronoun use, and inflections. Students have sufficient vocabulary to communicate in non-academic situations and most academic ones. They can engage in extended discussions. They can often use language to connect, tell, and expand on a topic; and can begin to use it to reason. They are fluent but may still hesitate in spontaneous communicative situations.

Level 3 (Intermediate). Students entering proficiency Level 3, the Intermediate level, are no longer wholly dependent on practiced, memorized, or formulaic language. They restructure learned language to communicate on a range of subjects. Their speech is still marked by errors in modality, tense, agreement, pronoun use, and inflections. These errors seldom interfere with communication in simple sentences, but do interfere in complex constructions. Intermediate level students are limited in vocabulary, especially academic vocabulary. They can retell, describe, narrate, question, and give instructions, although they lack fluidity and fluency when not using practiced or formulaic language. They often use language to connect, tell, and sometimes to expand on a known topic.

Level 2 (Beginning). Students who are just entering proficiency Level 2, the Beginning level, predominantly use formulaic patterns and memorized phrases. When they deviate from formulaic language, their speech imitates telegraphic language due to the omission of some meaningful linguistic components. Their language is also marked by the lack of tense, number, and agreement. They may use some very simple transitional markers, usually “and” to link ideas. They rely on schemata in L1. Their school-social vocabulary is limited to key words, and they have little or no academic vocabulary. They respond to questions usually with one- or two-word answers. They can connect and tell on a known topic.

Level 1 (Pre-functional). Students at this level may say or repeat common phrases, words, and formulaic language. They may be able to provide some basic information in response to requests and questions. They can ask one- or two-word questions without regard to structure and intonation.

Performance Level Definitions – Grades K–2 (cont’d)

WRITING

Level 5 (Fully English Proficient). Students at this level participate in writing activities with no teacher support. They write across all text types. They edit for sentence-level structure, spelling, and mechanics and revise for content, organization, and vocabulary. They can use complex sentence structures, with some errors, and can edit for syntax and grammar. They have a range of nonacademic and academic vocabulary that allows for precision, and they begin to use nuanced and alternative word meanings. They employ subtleties for different audiences and purposes. They can use appropriate writing conventions with some errors that do not affect comprehensibility.

Level 4 (Advanced). Students at this level participate in writing activities with minimal teacher support. They are able to restructure in writing the language they know to meet the creative demands of most social and academic situations. They can write mostly coherent, unified, and appropriately sequenced sentences. They use devices to connect ideas logically. They use a range of grammatical structures and can switch appropriately from one tense to another as required by the time frame of their text. They make some errors in modality, tense, agreement, pronoun use, and inflections. Students have a strong BICS vocabulary and a functional academic vocabulary that allows them to participate meaningfully in content classes. They write using all text types, at a developmentally appropriate level. They edit for sentence-level structure, spelling, and mechanics and revise for content, organization, and vocabulary.

Level 3 (Intermediate). Students at this level participate in writing activities with some teacher support. They can write simple and compound sentences and are beginning to write with phrases. They use simple tenses, number, and agreement, with random errors. They use transition words to link sentences and order these in a developmentally appropriate manner. They begin to edit for sentence-level structure, spelling, and mechanics, and revise for content, organization, and vocabulary, usually with the support of the teacher. They have a good range of BICS vocabulary and are beginning to use more academic content-specific words. They write mostly descriptive, expository, procedural, and narrative text. Their writing is less dependent on visual supports, shared experiences, and scaffolding.

Level 2 (Beginning). Students at this level participate in writing activities by drawing pictures or dictating words. They are able to write connected words and short telegraphic sentences. They are able to revise or edit their writing with teacher support. Their writing is marked by the lack of tense, number, and agreement. They may use some simple transitional markers, usually “and” to link ideas. Their vocabulary reflects what they can say orally. They make frequent errors in mechanics, such as punctuation and capitalization. They write mostly descriptive, expository, and procedural text. Their writing is most effective when supported by a visual, a shared experience, or scaffolding.

Level 1 (Pre-functional). Students at this level participate in writing activities by drawing pictures. They may be able to copy letters or form them from memory and may be able to copy some words. They can imitate the act of writing (e.g., scribbling); however, their text does not transmit a message. They may attempt to apply some writing conventions but do so inappropriately or do so correctly only when copying.

Performance Level Definitions – Grades 3–12

READING

Level 5 (Fully English Proficient). Students at this level understand the range of texts available to minimally proficient native English speakers, including literary and academic genres and texts from school-social settings. They understand main ideas and can extract precise and detailed information from a range of texts on familiar and unfamiliar topics in a number of genres comparable to a minimally proficient native English reader at the same grade level. They often successfully interpret text on the basis of understanding its purpose. They often successfully understand and can evaluate multiple perspectives of meaning. They understand complex structures of written English and have a broad range of vocabulary and idioms relating to both content areas and school-social environments.

Level 4 (Advanced). Students at this level understand most nonacademic and nontechnical texts appropriate for grade level. They understand many content area texts, mostly on familiar topics and approaching grade level. They understand excerpts from literature. They understand most written directions. They understand main ideas of a broad range of texts, especially when below grade level but also approaching grade level. They can begin to interpret text on the basis of understanding its purpose. They understand significant relevant details and can make subtle extrapolations of extended narratives or presentations on familiar academic topics. They understand sophisticated writer perspectives. They understand most of the basic language forms of written English and are beginning to develop understanding of more complex structures. They understand a wide range of vocabulary and idioms, especially of school-social environments, and are beginning to develop a wide range of technical vocabulary related to content areas.

Level 3 (Intermediate). Students at this level understand many authentic narrative and descriptive texts, especially when below grade level but with less complete comprehension for such texts on grade level. They understand content-area texts with familiar content, mostly when below grade level. They understand excerpts from literature, especially when below grade level. They understand simple written directions as well as some more complexly expressed directions. They understand main ideas of narrative and descriptive texts and some of the main points of expository and persuasive texts when they deal with areas of personal interest or topic familiarity. They begin to understand text purpose. They can understand some supporting ideas of expository and persuasive texts when dealing with areas of special interest. They understand some explicitly expressed points of view of writer and are able to draw simple conclusions. They understand frequently used verb tenses and word-order patterns in simple sentences. They understand a range of vocabulary and some idioms, mostly related to school-social environments, and have some key vocabulary from content areas.

Level 2 (Beginning). Students at this level understand short and simple authentic texts for informative or social purposes (e.g., general public statements, environmental texts, and formulaic messages). They have some understanding of short narrative texts or trade books, mostly when below grade level. They begin to understand some straightforward written directions. They understand main ideas and can identify a few explicit supporting ideas of simple, authentic, informative, and narrative materials when they contain simple language structures or rely heavily on visual cues or some prior experience with topic. They have some limited understanding text purpose. They are unable to extrapolate from text unless related to very basic ideas. They understand simple basic grammatical structures of written English in the school-social environment. They understand simple, basic, everyday vocabulary of the school environment and common everyday activities.

Level 1 (Pre-functional). Students at this level may identify isolated words and key phrases and cognates, especially when highly contextualized. They may understand some high-frequency, simple written directions, especially when highly contextualized. They are unable to identify any ideas intended by the writer of text or use limited knowledge of vocabulary and structural patterns to identify communicative intent of text or part of text. They do not understand how words, morphemes, and word order convey meaning in English.

Performance Level Definitions – Grades 3–12 (cont'd)

LISTENING

Level 5 (Fully English Proficient). Students at this level understand a significant amount of grade-level appropriate content-area and school-social speech. They understand the main ideas as well as relevant details and often subtle nuances of meaning of extended discussions or presentations on a range of familiar and unfamiliar topics comparable to a minimally proficient native English speaker at the same grade level. They are capable of making interpretations of what they listen to on the basis of understanding the speaker's purpose. They understand most of the complex structures of spoken English relative to their grade level. They have a broad range of vocabulary, including idiomatic language, relating to both content areas and school-social environments.

Level 4 (Advanced). Students at this level understand speech in most school-social settings and understand main ideas and some key supporting ideas in content-area settings. They understand multistep directions. They understand main ideas and significant relevant details of extended discussions or presentations on familiar and relevant academic topics. They can interpret text on the basis of understanding the purpose of text when it is on a familiar topic. They understand and are able to make subtle extrapolations from sophisticated speaker perspectives. They understand most of the basic language forms of spoken English and are beginning to develop understanding of more complex structures. They understand a wide range of vocabulary and idioms, especially of school-social environments, and are beginning to develop a wide range of technical vocabulary related to content areas.

Level 3 (Intermediate). Students at this level understand main ideas in short conversations on general school-social topics and frequently demonstrate general understanding of short messages or texts as well as longer conversations in familiar communicative situations and in academic content areas. They frequently demonstrate detailed understanding of short, discrete expressions but not of longer conversations and messages. They understand single-step and some multistep directions. They can begin to interpret text on the basis of understanding its purpose. They understand some explicitly expressed points of view and can draw simple conclusions. They understand frequently used verb tenses and word-order patterns in simple sentences. They understand a range of vocabulary and some idioms, mostly related to school-social environments, and have some key vocabulary from content areas.

Level 2 (Beginning). Students at this level understand simple and short statements, questions, and messages on familiar topics in school-social settings, and usually understand the main idea of simple messages and conversations. They can understand most common or critical information in the classroom but may identify and understand only key words, phrases, and cognates in content-area settings. They begin to understand straightforward, single-step directions and speaker's purpose. They have limited understanding of details and only of those that are explicitly stated and that support simple, straightforward messages or presentations. They are unable to extrapolate from text unless related to very basic ideas. They understand simple, basic grammatical structures and simple, basic, everyday vocabulary of spoken English in the school environment and common everyday activities.

Level 1 (Pre-functional). Students at this level may understand some common words or key phrases, especially when highly contextualized or when cognates. They may understand some high-frequency single-word or single-phrase directions, again, when highly contextualized. They generally are unable to use their limited knowledge of simple structural patterns to identify the communicative intent of the speaker.

Performance Level Definitions – Grades 3–12 (cont'd)

SPEAKING

Level 5 (Fully English Proficient). Students who are ready to enter Level 5, Fully English Proficient, can supply coherent, unified, and appropriately sequenced responses to an interlocutor. They use a variety of devices to connect ideas logically. They understand and can use a range of complex and simple grammatical structures as appropriate for topic and type of discourse. Their grammar and vocabulary is comparable to that of a minimally proficient native English speaker—grammar errors very seldom impede communication and their range of school-social and technical vocabulary allows a precision of speech comparable to a minimally proficient native English speaker. They infrequently but effectively use circumlocution. They can understand and use a variety of idiomatic phrases. They can effectively engage in non-interactive speech. They can use language effectively to connect, tell, expand, and reason. They show flexibility, creativity, and spontaneity in speech in a variety of contexts. Their pronunciation patterns (including stress and intonation) may be influenced by L1 but seldom interfere with communication.

Level 4 (Advanced). Students entering proficiency Level 4, the Advanced level, can supply mostly coherent, unified, and appropriately sequenced responses to an interlocutor. They use some devices to connect ideas logically and they use a range of grammatical structures. They make errors in modality, tense, agreement, pronoun use, and inflections, but these errors usually do not interfere with communication. Students have sufficient vocabulary to communicate in non-academic situations and some academic and technical vocabulary. They use circumlocutions and can appropriately use some idiomatic phrases. They can engage in extended discussions. They can often use language to connect, tell, and expand, and can begin to use it to reason. Their flexibility, creativity, and spontaneity are sometimes adequate for the communicative situation. Their pronunciation occasionally interferes with communication.

Level 3 (Intermediate). Students entering proficiency Level 3, the Intermediate level, display some use of discourse features but mainly rely on familiar, discrete utterances. They rely on simple transitional markers and use common, straightforward grammatical structures. They make errors in modality, tense, agreement, pronoun use, and inflections. These errors seldom interfere with communication in simple sentences but do interfere in complex constructions or when talking about academic issues. Intermediate level students are limited in vocabulary, especially academic and technical vocabulary. They use repetition; everyday, imprecise words; and code-switching to sustain conversations. They begin to use idiomatic expressions. They can retell, describe, narrate, question, and give simple, concrete instructions. They can often use language to connect and tell and sometimes to expand. They have some creativity and flexibility but often repeat themselves and hesitate. Their pronunciation patterns frequently interfere with communication.

Level 2 (Beginning). Students who are just entering proficiency Level 2, the Beginning level, use predominantly formulaic patterns in speech without regard to their connectivity. They may use some very simple transitional markers. They predominantly use formulaic patterns and memorized phrases, relying on schemata in L1. Their word order is frequently inappropriate, and frequent grammatical mistakes impede communication. Their school-social vocabulary is limited to key words; they have little or no technical vocabulary. They rely on survival vocabulary (needs and wants) and vocabulary provided by interlocutors. They may be able to name or list and can sometimes use language to connect or tell. Their limited vocabulary and knowledge of English structures impedes flexibility.

Level 1 (Pre-functional). Students in proficiency Level 1 are not yet at a functional level in English. They may repeat common phrases with very simple structures; be able to say a few, common, everyday words; and may be able to provide some basic information in response to requests.

Performance Level Definitions – Grades 3–12 (cont'd)

WRITING

Level 5 (Fully English Proficient). Students at this level demonstrate almost completely appropriate use of discourse features such as transition phrases and word order. They can revise for content, organization, and vocabulary. They can use complex sentence structures, with some errors, and can edit for syntax and grammar. They have a range of technical and nonacademic vocabulary that allows for precision, and they begin to use nuanced and alternative word meanings. They employ subtleties for different audiences and purposes. They can use appropriate writing conventions with some circumlocutions and errors that do not affect comprehensibility. Finally, they can successfully compose narrative, descriptive, expository, and persuasive texts.

Level 4 (Advanced). Students at this level demonstrate mostly successful use of discourse features such as transition words and sentence order. They can revise for content, organization and vocabulary and show good control of the most frequently used grammatical structures, with errors. They can edit for sentence-level structure. They have sufficient vocabulary to express themselves with some circumlocutions, which are more frequent in academic contexts. Their tone indicates some awareness of audience. They can use appropriate writing conventions, with circumlocutions and errors that infrequently affect comprehensibility. Finally, they can successfully compose narrative and descriptive texts and they may be successful writing expository and persuasive texts.

Level 3 (Intermediate). Students at this level demonstrate some use of discourse features such as transition words and sentence order. They begin to revise for content, organization and vocabulary. They demonstrate comprehensible use of basic sentence structures, with errors, and can begin to edit for sentence-level structure. They use everyday vocabulary but know very few content-specific words. There is some variation in their register, voice, and tone. They make frequent mechanical errors, particularly when expressing complex thoughts or technical ideas. Finally, students can compose narrative and some descriptive texts and can begin to write expository and persuasive texts.

Level 2 (Beginning). Students at this level may or may not use some basic rhetorical features, such as ordering sentences appropriately and using simple cohesive devices. They are unlikely to revise their writing spontaneously. Their writing is limited to typical, present-tense, subject-verb-object sentences or phrases and is likely to be repetitive. They edit only with explicit support and direction and have a limited vocabulary. They make frequent errors in mechanics, which is characteristic and expected. Their text range is limited to narrative or simple descriptive.

Level 1 (Pre-functional). Students at this level are not yet functional in English. They might be able to copy letters or form them from memory and might be able to write words; however, their text does not transmit a coherent message. They do not use discourse features in their writing. There is no evidence of appropriate text structure, and sentence-level structure is predominantly inappropriate. They may attempt to apply some writing conventions but do so inappropriately or do so correctly only when copying.

Performance Level Definitions – Grades 3–12 (cont'd)

COMPREHENSION

Level 5 (Fully English Proficient). Students at this level understand the range of texts available to minimally proficient native English speakers on content-area and non-content-area topics. They understand main ideas, relevant details, and often subtle nuances of meaning of a range of texts on familiar and unfamiliar topics, comparable to a minimally proficient native English speaker at the same grade level. They often successfully interpret text on the basis of understanding its purpose. They understand and can evaluate multiple-writer or speaker perspectives. They understand complex structures of English and have a broad range of vocabulary relating to both content areas and school social environments.

Level 4 (Advanced). Students at this level understand most non-academic and non-technical texts appropriate for grade level. They understand main ideas and some key supporting ideas in content-area texts, mostly on familiar topics and approaching grade level. They understand most multistep directions. They understand main ideas and significant relevant details of a broad range of texts on familiar and relevant academic topics, especially when below grade level but also approaching grade level. They can begin to interpret text on the basis of understanding its purpose. They understand and are able to make subtle extrapolations from sophisticated writer and speaker perspectives. They understand most of the basic language forms of English and are beginning to develop understanding of more complex structures.

Level 3 (Intermediate). Students at this level understand short simple texts on general school-related topics, especially when below grade level, but their comprehension is less complete for such texts on grade level. They understand content-area texts with familiar content, mostly when below grade level. They understand single-step and some multistep directions. They understand main ideas and some supporting ideas of short, simple texts when they deal with areas of personal interest in familiar communicative situations and academic content areas. They begin to understand text purpose. They can understand some complex text types, especially when dealing with areas of special interest. They understand some explicitly expressed points of view and are able to draw simple conclusions. They understand frequently used verb tenses and word-order patterns in simple sentences. They understand short, simple text on familiar topics, especially when containing formulaic language.

Level 2 (Beginning). Students at this level may only identify and understand key words, phrases, and cognates in content area settings. They can begin to follow straightforward, single-step directions. They usually understand main ideas of simple texts when they contain simple language structures and/or rely heavily on visual cues and/or some prior experience with topic. They have some limited understanding of purpose of text and have limited understanding of details and only of those that are explicitly stated. They are unable to extrapolate from text unless related to very basic ideas. They understand simple, basic grammatical structures of English in the school environment. They understand simple, basic, everyday vocabulary of the school environment and common everyday activities.

Level 1 (Pre-functional). Students at this level may recognize some common words or key phrases, especially when they are highly contextualized or when they are cognates. They may understand some high frequency single-word or phrase directions, again, when highly contextualized. They are unable to identify any ideas or use their limited knowledge of simple structural patterns or vocabulary to identify writer or speaker communicative intent.

Appendix C

2015 Sample Score Reports

Sample Individual Student Report

English Language Development Assessment

Score Report for: LASTNAME, FIRSTNAME **ID:** 11111111111
Birth Date: 01/01/2007 **Cluster:** 3-5
District/Code: ABC Public Schools / 999999000 **Grade:** 03
School/Code: ABC Elementary School / 999999999 **Test Date:** Spring 2015



Listening		Speaking		Reading		Writing		Comprehension*	Composite**
Score	Level	Score	Level	Score	Level	Score	Level	Level	Level
612	3	809	5	601	3	506	2	3	2

RT = Zero Score

*Comprehension is derived from Listening and Reading (available only for students with valid scores on both tests).

**Composite is derived from Listening, Speaking, Reading and Writing (available only for students with valid scores on all four tests).

The English Language Development Assessment (ELDA) measures the English language skills of students whose first language is other than English. There are scores for Listening, Speaking, Reading, Writing, Comprehension and Composite.

Explanation of Levels

Level 1 - Pre-functional indicates that the student who is limited English proficient is:

- Beginning to understand short utterances
- Beginning to use gestures and simple words to communicate
- Beginning to understand simple printed material
- Beginning to develop communicative writing skills

Level 2 - Beginning indicates that the student who is limited English proficient can:

- Understand simple statements, directions, and questions
- Use appropriate strategies to initiate and respond to simple conversation
- Understand the general message of basic reading passages
- Compose short informative passages on familiar topics

Level 3 - Intermediate indicates that the student who is limited English proficient can:

- Understand standard speech delivered in school and social settings
- Communicate orally with some hesitation
- Understand descriptive material within familiar contexts and some complex narratives
- Write simple texts and short reports

Level 4 - Advanced indicates that the student who is limited English proficient can:

- Identify the main ideas and relevant details of discussions or presentations on a wide range of topics
- Actively engage in most communicative situations familiar or unfamiliar
- Understand the context of most text in academic areas with support
- Write multi-paragraph essays, journal entries, personal/business letters, and creative texts in an organized fashion with some errors

Level 5 - Full English Proficiency indicates that the student who is limited English proficient can:

- Understand and identify the main ideas and relevant details of extended discussion or presentations on familiar and unfamiliar topics
- Produce fluent and accurate language
- Use reading strategies the same as their native English-speaking peers to derive meaning from a wide range of both social and academic texts
- Write fluently using language structures, technical vocabulary, and appropriate writing conventions with some circumlocutions

English Language Development Assessment
Student Roster Report



State: NE
 District/Code: ABC Public Schools / 999999000
 School/Code: ABC Elementary School / 999999999
 Cluster: 3-5
 Grade: 03
 Test Date: Spring 2015

Student Name	Student ID	Listening		Speaking		Reading		Writing		Comprehension*		Composite**	
		Score	Level	Score	Level	Score	Level	Score	Level	Level	Level	Level	Level
LAST, FIRST	1111111111	442	1	604	3	377	1	357	1	1	1	1	1
LAST, FIRST	2222222222	460	2	587	3	338	1	330	1	1	1	1	1
LAST, FIRST	3333333333	387	1	940	5	413	1	506	2	1	1	1	1
LAST, FIRST	4444444444	594	3	758	4	463	2	401	1	2	2	1	1
LAST, FIRST	5555555555	649	4	890	5	641	3	557	2	3	3	2	2
LAST, FIRST	6666666666	649	4	643	3	689	4	635	3	4	4	3	3
LAST, FIRST	7777777777	612	3	890	5	641	3	576	2	3	3	2	2
LAST, FIRST	8888888888	526	2	758	4	480	2	594	3	2	2	2	2

Level = Proficiency Level (1 = Pre-Functional; 5 = Fully English Proficient; RT = Zero Score)
 *Comprehension is derived from Listening and Reading (available only for students with valid scores on both tests).
 **Composite is derived from Listening, Speaking, Reading and Writing (available only for students with valid scores on all four tests).

English Language Development Assessment
District Summary Report



State: NE
 District/Code: ABC Public Schools / 999999000
 Cluster: 3-5
 Grade: 03
 Test Date: Spring 2015
 Score Range: 426-722

Proficiency Level	Total Students	Listening				Speaking				Reading				Writing				Comprehension*				Composite**			
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%		
5	N	14		14		14		14		14		14		14		14		14		14		14			
	%	1		6		43		0		0		0		0		0		0		0		0			
4	N	2		3		3		2		1		2		1		2		2		2		0			
	%	14		21		21		14		7		14		7		14		14		14		0			
3	N	6		5		5		3		3		3		3		4		4		4		4			
	%	43		36		36		21		21		21		21		29		29		29		29			
2	N	2		0		0		4		6		4		6		3		3		3		4			
	%	14		0		0		29		43		29		43		21		21		21		29			
1	N	3		0		0		5		4		5		4		5		5		5		6			
	%	21		0		0		36		29		36		29		36		36		36		43			
Mean Scale Score	District	564				761				518				520				542				591			
Standard Deviation	District	117				133				134				109				119				104			
Lowest Score	District	345				587				316				330				331				426			
Highest Score	District	769				940				717				684				674				722			

*Comprehension is derived from Listening and Reading (available only for students with valid scores on both tests).
 **Composite is derived from Listening, Speaking, Reading and Writing (available only for students with valid scores on all four tests).

Sample District Demographic Report

English Language Development Assessment
District Demographic Report

State: NE
District/Code: ABC Public Schools / 999999000
Cluster: 3-5
Grade: 03
Test Date: Spring 2015



	# Valid Scores	Listening		Speaking		Reading		Writing		Comprehension*		Composite**	
		% FEP	Mean Scale Score	% FEP	Mean Scale Score	% FEP	Mean Scale Score	% FEP	Mean Scale Score	% FEP	Mean Scale Score	% FEP	Mean Scale Score
Total Students	14	7	564	43	761	0	518	0	520	0	542	0	591
General Students	14	7	564	43	761	0	518	0	520	0	542	0	591
Section 504 Students	0	0	0	0	0	0	0	0	0	0	0	0	0
Gender													
Female	5	0	561	40	784	0	554	0	574	0	557	0	618
Male	9	11	566	44	747	0	499	0	490	0	533	0	575
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Ethnicity													
American Indian / Alaska Native	0	0	0	0	0	0	0	0	0	0	0	0	0
Asian	0	0	0	0	0	0	0	0	0	0	0	0	0
Native Hawaiian / Pacific Islander	0	0	0	0	0	0	0	0	0	0	0	0	0
Black or African American	0	0	0	0	0	0	0	0	0	0	0	0	0
Hispanic	14	7	564	43	761	0	518	0	520	0	542	0	591
White	0	0	0	0	0	0	0	0	0	0	0	0	0
Multiple / Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Length of ESL													
< 1 Year	0	0	0	0	0	0	0	0	0	0	0	0	0
1 - 2 Years	0	0	0	0	0	0	0	0	0	0	0	0	0
2 - 3 Years	0	0	0	0	0	0	0	0	0	0	0	0	0
3 - 4 Years	1	0	576	0	604	0	430	0	490	0	503	0	525
4 - 5 Years	1	0	630	0	691	0	546	0	557	0	588	0	606
5 + Years	0	0	0	0	0	0	0	0	0	0	0	0	0
First Language													
Spanish	14	7	564	43	761	0	518	0	520	0	542	0	591
Arabic	0	0	0	0	0	0	0	0	0	0	0	0	0
Vietnamese	0	0	0	0	0	0	0	0	0	0	0	0	0
Karen	0	0	0	0	0	0	0	0	0	0	0	0	0
Nuer	0	0	0	0	0	0	0	0	0	0	0	0	0
Somali	0	0	0	0	0	0	0	0	0	0	0	0	0
Chinese	0	0	0	0	0	0	0	0	0	0	0	0	0
Russian	0	0	0	0	0	0	0	0	0	0	0	0	0
Ukrainian	0	0	0	0	0	0	0	0	0	0	0	0	0
French	0	0	0	0	0	0	0	0	0	0	0	0	0
Kurdish	0	0	0	0	0	0	0	0	0	0	0	0	0
Dinka	0	0	0	0	0	0	0	0	0	0	0	0	0
Other / Missing	0	0	0	0	0	0	0	0	0	0	0	0	0

*Comprehension is derived from Listening and Reading (available only for students with valid scores on both tests).

**Composite is derived from Listening, Speaking, Reading and Writing (available only for students with valid scores on all four tests).

This report only includes students with a Composite score.

Appendix D

2015 Operational Item Statistics

Table D.1
2015 Operational Item Statistics
Cluster 1 Kindergarten Reading Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.37	0.84	0.70	3334	4.05	10.74	29.12	55.94	0.15
2	2.40	0.85	0.64	3334	4.86	9.09	26.30	59.54	0.21
3	1.86	0.95	0.76	3334	9.15	25.04	35.99	29.69	0.12
4	1.50	1.06	0.78	3334	20.64	30.47	26.27	22.44	0.18
5	1.97	0.95	0.78	3334	7.41	23.94	32.60	35.90	0.15
6	2.56	0.74	0.65	3334	2.10	8.34	20.40	69.02	0.15
7	1.68	0.89	0.78	3334	8.52	34.40	37.19	19.65	0.21
8	1.34	0.96	0.76	3334	21.75	35.12	30.02	12.93	0.18
9	1.10	0.89	0.67	3334	28.22	39.02	26.21	6.33	0.21
10	1.93	0.91	0.70	3334	6.48	25.19	36.71	31.37	0.24
11	1.81	0.91	0.81	3334	8.55	26.81	39.41	24.99	0.24
12	1.51	0.96	0.81	3334	14.94	36.35	30.47	18.03	0.21
13	1.29	0.94	0.79	3334	22.38	36.77	29.75	10.89	0.21
14	1.38	0.95	0.77	3334	18.39	40.04	26.27	15.00	0.30

Table D.2
2015 Operational Item Statistics
Cluster 1 Kindergarten Listening Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.06	0.83	0.77	3334	2.88	22.77	39.86	34.40	0.09
2	1.87	0.79	0.83	3334	3.36	27.89	46.46	22.20	0.09
3	1.54	0.85	0.85	3334	10.65	37.34	39.32	12.57	0.12
4	1.53	0.84	0.85	3334	10.83	37.46	39.65	11.97	0.09
5	2.03	0.85	0.81	3334	3.57	23.61	38.93	33.77	0.12
6	1.80	0.84	0.84	3334	5.07	31.13	41.51	22.11	0.15
7	1.50	0.86	0.82	3334	12.75	35.81	39.44	11.76	0.24

Table D.3
2015 Operational Item Statistics
Cluster 1 Kindergarten Writing Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	1.56	0.87	0.75	3334	8.22	44.27	30.20	17.04	0.27
2	2.08	0.90	0.79	3334	4.50	22.20	33.56	39.47	0.27
3	1.39	0.87	0.78	3334	16.35	37.13	37.16	9.03	0.33
4	1.55	0.99	0.83	3334	16.65	31.13	32.60	19.44	0.18
5	1.30	0.82	0.80	3334	18.66	35.66	41.45	3.96	0.27
6	1.84	0.77	0.72	3334	4.26	25.58	51.23	18.75	0.18
7	1.50	0.75	0.72	3334	5.46	48.02	36.62	9.66	0.24
8	1.62	0.91	0.84	3334	12.93	28.82	41.24	16.77	0.24
9	1.65	0.92	0.79	3334	10.83	32.63	36.38	19.80	0.36

Table D.4
2015 Operational Item Statistics
Cluster 1 Kindergarten Speaking Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.21	0.78	0.81	3334	2.70	13.80	42.59	40.79	0.12
2	2.01	0.82	0.84	3334	4.20	19.68	46.04	29.87	0.21
3	1.99	0.81	0.84	3334	3.33	22.86	44.39	29.27	0.15
4	2.06	0.83	0.83	3334	3.45	21.15	40.70	34.58	0.12
5	1.93	0.85	0.84	3334	4.83	24.93	42.62	27.50	0.12
6	1.79	0.82	0.85	3334	6.81	25.22	49.64	18.15	0.18
7	1.58	0.86	0.78	3334	12.30	29.48	45.47	12.60	0.15
8	1.63	0.85	0.82	3334	10.14	30.35	45.08	14.04	0.39

Table D.5
2015 Operational Item Statistics
Cluster 2 Grades 1–2 Reading Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.75	0.57	0.56	5541	0.96	3.47	14.11	81.23	0.22
2	1.90	0.86	0.79	5541	4.58	28.46	38.55	28.26	0.14
3	1.81	0.83	0.81	5541	5.14	30.16	43.11	21.42	0.16
4	1.88	0.96	0.82	5541	9.67	22.79	36.85	30.50	0.18
5	2.00	0.82	0.81	5541	3.59	21.78	45.08	29.20	0.32
6	2.34	0.82	0.74	5541	2.96	13.45	29.80	53.64	0.16
7	1.82	0.85	0.80	5541	4.98	31.17	40.55	23.14	0.16
8	1.80	0.83	0.83	5541	5.54	29.20	44.29	20.79	0.18
9	1.53	0.88	0.81	5541	13.81	31.37	42.34	12.31	0.18
10	2.09	0.89	0.75	5541	5.40	18.52	37.01	38.78	0.29
11	2.33	0.80	0.79	5541	2.78	12.04	33.95	51.06	0.18
12	1.87	0.86	0.84	5541	5.99	25.63	42.83	25.28	0.27
13	1.60	0.86	0.81	5541	10.41	32.58	42.54	14.19	0.29
14	1.82	0.90	0.83	5541	8.27	24.65	42.56	24.20	0.32

Table D.6
2015 Operational Item Statistics
Cluster 2 Grades 1–2 Listening Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.31	0.76	0.76	5541	1.34	13.55	37.14	47.79	0.18
2	2.17	0.73	0.82	5541	1.59	14.28	48.98	34.99	0.16
3	1.90	0.81	0.85	5541	4.85	23.10	48.49	23.28	0.27
4	1.87	0.81	0.85	5541	4.67	25.48	47.59	22.05	0.20
5	2.34	0.77	0.79	5541	1.57	12.98	34.56	50.69	0.20
6	2.05	0.81	0.81	5541	2.53	21.71	43.22	32.21	0.32
7	1.85	0.80	0.81	5541	5.38	23.80	50.42	20.18	0.22

Table D.7
2015 Operational Item Statistics
Cluster 2 Grades 1–2 Writing Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.17	0.82	0.74	5541	2.24	18.84	37.20	41.38	0.34
2	2.48	0.77	0.80	5541	2.42	9.02	25.88	62.30	0.38
3	2.25	0.83	0.75	5541	3.75	12.40	37.81	45.62	0.42
4	1.78	0.69	0.77	5541	5.68	18.81	66.20	9.01	0.31
5	1.68	0.74	0.78	5541	6.79	27.03	56.87	8.99	0.31
6	2.30	0.70	0.67	5541	1.06	9.82	45.89	42.83	0.40
7	1.92	0.76	0.71	5541	3.05	23.19	51.24	22.20	0.32
8	1.95	0.83	0.78	5541	4.51	22.18	45.68	27.27	0.36
9	1.95	0.83	0.79	5541	4.75	21.84	45.71	27.29	0.42

Table D.8
2015 Operational Item Statistics
Cluster 2 Grades 1–2 Speaking Inventory

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	omit
1	2.47	0.71	0.77	5541	1.53	7.31	33.01	57.90	0.25
2	2.32	0.77	0.79	5541	2.02	11.84	38.01	47.90	0.23
3	2.33	0.76	0.82	5541	1.86	11.15	37.79	48.87	0.32
4	2.19	0.73	0.82	5541	2.49	10.68	51.56	35.03	0.23
5	2.27	0.77	0.83	5541	2.22	12.47	40.61	44.47	0.22
6	2.15	0.77	0.83	5541	2.35	15.29	46.06	36.02	0.29
7	1.81	0.82	0.76	5541	7.33	21.64	52.57	18.21	0.25
8	1.86	0.82	0.79	5541	6.95	19.74	52.99	20.03	0.29

Table D.9
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Reading

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.85	0.49	5254	7.37	84.94	5.60	1.83	0.25	0.02
2	0.93	0.44	5254	93.15	2.44	1.20	2.55	0.61	0.06
3	0.94	0.35	5254	1.43	1.81	94.35	1.73	0.69	0.00
4	0.77	0.42	5254	8.36	10.68	3.08	77.22	0.65	0.02
5	0.65	0.38	5254	17.85	9.35	64.85	7.19	0.74	0.02
6	0.80	0.53	5254	7.44	5.77	80.32	5.61	0.86	0.00
7	0.84	0.46	5254	5.61	84.07	4.30	5.50	0.51	0.00
8	0.61	0.43	5254	22.14	10.98	60.53	5.50	0.86	0.00
9	0.59	0.37	5254	8.62	9.69	58.56	22.23	0.89	0.00
10	0.74	0.43	5254	5.35	73.66	15.76	4.26	0.97	0.00
11	0.79	0.45	5254	78.82	7.14	7.42	5.48	1.12	0.02
12	0.69	0.36	5254	68.75	11.95	6.93	11.67	0.70	0.00
13	0.77	0.54	5254	77.08	10.24	4.15	7.42	1.03	0.08
14	0.80	0.58	5254	6.57	5.71	6.59	79.92	1.20	0.02
15	0.45	0.30	5254	8.39	45.34	34.77	10.26	1.24	0.00
16	0.84	0.45	5254	7.80	84.24	2.72	3.96	1.18	0.10
17	0.86	0.59	5254	6.20	4.05	2.76	85.78	1.20	0.00
18	0.86	0.58	5254	6.03	4.00	3.25	85.55	1.14	0.02
19	0.81	0.56	5254	80.99	5.35	8.22	4.30	1.12	0.02
20	0.81	0.50	5254	6.05	7.69	80.99	4.05	1.20	0.02
21	0.86	0.53	5254	85.92	3.94	4.51	4.36	1.24	0.04
22	0.63	0.40	5254	62.56	11.48	11.42	13.30	1.20	0.04
23	0.44	0.32	5254	17.80	23.94	43.93	13.09	1.22	0.02
24	0.55	0.33	5254	15.13	22.65	6.09	54.99	1.14	0.00
25	0.49	0.35	5254	12.83	7.86	28.55	49.49	1.20	0.08
26	0.57	0.41	5254	57.08	12.26	10.18	19.11	1.29	0.08
27	0.63	0.50	5254	15.47	11.91	62.83	8.36	1.37	0.06
28	0.65	0.51	5254	65.15	15.44	9.04	8.91	1.43	0.04
29	0.58	0.39	5254	58.41	7.77	17.03	15.34	1.45	0.00
30	0.57	0.44	5254	26.38	56.97	10.01	5.12	1.52	0.00
31	0.82	0.48	5254	81.79	3.69	7.02	5.77	1.67	0.06
32	0.56	0.40	5254	23.98	11.84	56.26	5.90	1.96	0.06
33	0.73	0.46	5254	11.90	72.93	3.37	9.88	1.85	0.08
34	0.68	0.48	5254	6.62	68.18	13.09	10.28	1.83	0.00
35	0.60	0.48	5254	59.84	14.07	7.99	16.22	1.88	0.00

Table D.10
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Listening

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.36	0.18	5254	31.18	8.96	35.57	23.85	0.44	0.00
2	0.77	0.42	5254	5.63	77.10	8.58	7.86	0.80	0.02
3	0.77	0.47	5254	5.54	4.82	12.09	76.84	0.72	0.00
4	0.85	0.30	5254	4.74	2.36	7.35	84.79	0.76	0.00
5	0.85	0.46	5254	84.55	4.61	3.18	7.18	0.44	0.06
6	0.83	0.51	5254	1.58	4.61	82.68	10.39	0.72	0.02
7	0.86	0.52	5254	5.88	86.11	4.15	3.14	0.70	0.02
8	0.58	0.30	5254	58.34	29.33	3.22	8.62	0.48	0.02
9	0.80	0.43	5254	11.91	3.35	4.00	80.00	0.70	0.04
10	0.92	0.42	5254	1.12	4.09	91.66	2.36	0.74	0.02
11	0.86	0.48	5254	2.93	85.95	5.73	4.57	0.80	0.02
12	0.59	0.33	5254	16.84	8.60	58.72	15.00	0.82	0.02
13	0.70	0.30	5254	6.36	15.42	7.56	69.83	0.82	0.02
14	0.91	0.45	5254	3.58	90.98	2.17	2.47	0.80	0.00
15	0.74	0.38	5254	4.15	15.36	73.96	5.69	0.84	0.00
16	0.75	0.45	5254	9.94	75.05	7.65	6.57	0.80	0.00
17	0.79	0.50	5254	79.25	7.00	7.04	5.88	0.78	0.04
18	0.74	0.42	5254	3.22	18.20	3.37	74.40	0.82	0.00
19	0.40	0.28	5254	15.66	3.31	40.08	40.08	0.86	0.00
20	0.58	0.38	5254	12.16	58.09	17.74	11.15	0.82	0.04
21	0.62	0.37	5254	5.98	20.12	61.88	11.12	0.89	0.02
22	0.87	0.41	5254	3.98	86.60	3.46	5.06	0.89	0.00
23	0.90	0.44	5254	3.31	2.66	90.39	2.65	0.97	0.02
24	0.71	0.30	5254	71.32	8.49	4.85	14.66	0.69	0.00
25	0.70	0.37	5254	3.43	69.78	23.26	2.61	0.93	0.00
26	0.39	0.13	5254	21.58	23.22	15.61	38.56	0.99	0.04
27	0.26	0.17	5254	32.81	26.19	20.97	18.90	1.05	0.08
28	0.58	0.34	5254	57.84	13.21	15.15	12.77	1.03	0.00
29	0.44	0.39	5254	20.08	14.31	20.88	43.59	1.08	0.06
30	0.39	0.27	5254	9.54	12.90	39.17	37.27	1.07	0.06
31	0.37	0.14	5254	29.60	36.79	17.99	14.50	1.07	0.06
32	0.61	0.36	5254	26.76	60.62	7.25	4.34	1.03	0.00
33	0.76	0.40	5254	10.83	5.88	76.04	6.15	1.07	0.04
34	0.86	0.52	5254	4.11	85.53	5.69	3.58	1.08	0.00
35	0.73	0.44	5254	3.18	73.32	15.63	6.70	1.18	0.00

Table D.11
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Writing (CR)

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	4
1	2.58	0.75	0.62	5254	3.71	4.68	21.81	69.79	
2	2.25	0.80	0.62	5254	4.47	9.29	43.01	43.22	
3	2.49	0.78	0.65	5254	4.42	4.70	28.34	62.54	
4	2.30	0.95	0.64	5254	5.37	10.24	40.83	35.78	7.78

Table D.12
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Writing (MC)

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.46	0.27	5254	23.58	46.38	15.00	14.43	0.61	0.00
2	0.71	0.40	5254	71.15	8.60	12.68	6.41	1.07	0.10
3	0.71	0.38	5254	9.00	71.36	10.64	7.82	1.14	0.04
4	0.59	0.33	5254	5.06	59.40	30.55	3.65	1.31	0.02
5	0.77	0.45	5254	6.66	7.08	77.20	7.77	1.26	0.04
6	0.55	0.44	5254	4.42	9.15	30.49	54.68	1.24	0.02
7	0.53	0.27	5254	53.01	8.26	15.80	21.66	1.24	0.04
8	0.50	0.23	5254	50.38	17.53	19.72	10.98	1.33	0.06
9	0.46	0.30	5254	14.01	20.65	46.48	17.36	1.48	0.02
10	0.40	0.22	5254	4.11	10.60	43.47	40.33	1.47	0.02
11	0.67	0.43	5254	8.68	67.17	13.36	9.33	1.45	0.02
12	0.61	0.31	5254	60.62	9.99	9.12	18.67	1.60	0.00
13	0.40	0.30	5254	5.81	11.76	40.94	40.29	1.20	0.00
14	0.37	0.15	5254	19.51	23.85	37.27	17.68	1.62	0.08
15	0.41	0.21	5254	13.53	41.38	25.94	17.47	1.67	0.00

Table D.13a
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Speaking

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point			
					0	1	2	Omit
1	1.64	0.71	0.63	5254	7.54	9.95	76.78	5.73
2	1.84	0.47	0.74	5254	2.40	7.38	88.16	2.06
3	1.65	0.59	0.65	5254	3.81	22.92	70.86	2.42
4	1.16	0.73	0.52	5254	15.15	44.10	36.01	4.74
5	1.75	0.55	0.68	5254	3.14	13.08	81.12	2.66
6	1.89	0.41	0.73	5254	1.66	3.69	92.79	1.85
7	1.72	0.55	0.67	5254	2.91	18.12	76.89	2.07
8	1.46	0.66	0.64	5254	6.20	35.59	55.39	2.82
9	1.77	0.57	0.69	5254	4.09	8.45	84.18	3.25
10	1.83	0.48	0.73	5254	2.17	8.45	87.23	2.15
11	1.78	0.51	0.71	5254	2.19	13.46	82.34	2.02
12	1.48	0.67	0.65	5254	6.09	33.04	57.25	3.62

Table D.13b
2015 Operational Item Statistics
Cluster 3 Grades 3–5 Speaking Condition Codes

Item position	Percentage of respondents omitting item at each condition code					Total number of respondents omitting item
	A	B	C	D	No code	
1	3.32	59.46	20.94	12.62	3.66	301
2	4.62	42.60	26.86	13.88	12.04	108
3	3.14	42.52	25.98	18.12	10.24	127
4	2.40	51.80	25.30	13.66	6.82	249
5	2.86	42.86	22.86	20.72	10.72	140
6	1.04	35.06	31.96	14.44	17.52	97
7	9.18	38.54	30.28	8.26	13.76	109
8	2.70	46.62	22.30	17.56	10.82	148
9	3.50	43.86	20.46	23.40	8.78	171
10	6.20	45.14	24.78	10.62	13.28	113
11	4.72	41.50	27.36	12.26	14.16	106
12	3.16	48.94	22.10	17.36	8.42	190

Condition Codes:

- A** - inaudible/unclear
- B** - blank
- C** - refusal or excuse (“I don’t want to answer,” “I don’t know”)
- D** - off topic/off task/completely in language other than English

Table D.14
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Reading

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.84	0.48	2036	5.70	84.48	3.44	5.70	0.69	0.00
2	0.83	0.44	2036	83.01	5.35	8.60	1.82	1.18	0.05
3	0.38	0.26	2036	11.89	38.65	9.82	38.31	1.33	0.00
4	0.60	0.50	2036	13.85	10.46	60.17	14.05	1.47	0.00
5	0.86	0.55	2036	1.96	5.65	86.44	4.57	1.33	0.05
6	0.52	0.30	2036	33.69	7.02	5.80	52.06	1.42	0.00
7	0.55	0.23	2036	10.27	14.00	18.86	55.35	1.42	0.10
8	0.74	0.47	2036	13.56	73.77	4.52	7.12	1.03	0.00
9	0.51	0.42	2036	11.10	21.71	50.69	14.93	1.57	0.00
10	0.80	0.43	2036	9.04	79.52	3.29	6.43	1.72	0.00
11	0.69	0.52	2036	5.75	12.48	11.44	68.86	1.42	0.05
12	0.67	0.61	2036	6.04	11.30	67.49	13.26	1.82	0.10
13	0.78	0.46	2036	9.97	78.34	5.65	4.22	1.67	0.15
14	0.39	0.35	2036	37.33	12.33	39.34	8.74	2.16	0.10
15	0.47	0.47	2036	47.45	17.09	10.71	22.69	1.96	0.10
16	0.71	0.58	2036	71.02	11.74	8.60	6.58	2.01	0.05
17	0.76	0.59	2036	76.47	7.02	9.09	5.45	1.96	0.00
18	0.55	0.40	2036	13.16	54.91	14.29	15.91	1.67	0.05
19	0.77	0.56	2036	77.31	8.50	5.89	6.48	1.77	0.05
20	0.72	0.49	2036	5.50	71.56	11.10	9.92	1.92	0.00
21	0.75	0.65	2036	11.59	6.58	75.00	4.72	2.01	0.10
22	0.63	0.55	2036	17.63	12.62	62.67	5.01	2.06	0.00
23	0.63	0.52	2036	9.09	11.00	14.93	62.72	2.26	0.00
24	0.75	0.59	2036	5.16	5.40	74.75	12.72	1.96	0.00
25	0.59	0.44	2036	59.33	8.99	17.09	12.57	2.01	0.00
26	0.58	0.49	2036	58.50	20.58	13.02	5.80	2.11	0.00
27	0.71	0.53	2036	13.26	8.74	70.58	5.26	2.11	0.05
28	0.79	0.54	2036	7.56	78.73	6.93	4.57	2.11	0.10
29	0.59	0.59	2036	9.53	17.49	11.44	59.14	2.31	0.10
30	0.75	0.60	2036	8.84	5.35	75.39	8.06	2.21	0.15
31	0.81	0.62	2036	80.89	3.83	5.50	7.37	2.36	0.05
32	0.73	0.52	2036	9.48	73.23	9.97	4.91	2.36	0.05
33	0.68	0.50	2036	11.00	5.55	67.98	13.02	2.41	0.05
34	0.53	0.45	2036	13.75	10.22	20.73	52.70	2.60	0.00
35	0.58	0.49	2036	57.91	11.79	16.94	10.85	2.50	0.00

Table D.15
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Listening

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.73	0.40	2036	13.56	73.28	10.41	1.77	0.93	0.05
2	0.84	0.58	2036	4.27	4.37	83.99	5.84	1.52	0.00
3	0.72	0.45	2036	8.94	7.56	71.76	10.31	1.42	0.00
4	0.72	0.52	2036	72.45	14.10	7.96	4.03	1.47	0.00
5	0.63	0.47	2036	17.68	7.76	62.57	10.81	1.13	0.05
6	0.83	0.51	2036	5.30	7.22	82.51	3.54	1.38	0.05
7	0.82	0.65	2036	81.97	9.58	3.24	3.73	1.42	0.05
8	0.74	0.58	2036	11.94	2.95	10.61	73.53	0.98	0.00
9	0.80	0.53	2036	80.16	7.66	2.46	8.35	1.38	0.00
10	0.84	0.50	2036	83.60	4.67	2.06	8.15	1.47	0.05
11	0.86	0.55	2036	5.40	3.93	86.20	3.05	1.42	0.00
12	0.73	0.46	2036	6.24	72.84	10.90	8.50	1.47	0.05
13	0.65	0.47	2036	7.32	17.53	64.98	8.64	1.52	0.00
14	0.77	0.46	2036	77.16	7.76	3.44	10.12	1.47	0.05
15	0.76	0.42	2036	3.63	75.79	10.76	8.40	1.42	0.00
16	0.86	0.60	2036	5.01	85.90	3.49	4.13	1.47	0.00
17	0.77	0.49	2036	3.88	6.34	11.35	76.96	1.42	0.05
18	0.84	0.63	2036	5.84	5.60	83.84	3.24	1.47	0.00
19	0.84	0.62	2036	2.36	83.94	6.78	5.40	1.52	0.00
20	0.84	0.51	2036	5.16	83.69	5.45	4.47	1.13	0.10
21	0.63	0.34	2036	7.66	9.48	62.72	18.52	1.47	0.15
22	0.53	0.43	2036	17.63	10.85	17.44	52.55	1.47	0.05
23	0.51	0.39	2036	13.31	16.11	18.42	50.54	1.52	0.10
24	0.75	0.62	2036	6.53	12.82	75.29	4.22	1.03	0.10
25	0.64	0.49	2036	7.96	64.15	7.96	18.37	1.47	0.10
26	0.60	0.32	2036	59.72	20.04	7.76	10.46	1.96	0.05
27	0.69	0.32	2036	10.66	11.35	7.66	68.76	1.52	0.05
28	0.81	0.49	2036	5.60	80.50	7.07	5.21	1.57	0.05
29	0.56	0.46	2036	12.62	56.24	4.17	25.39	1.47	0.10
30	0.67	0.51	2036	11.49	15.23	4.76	67.04	1.42	0.05
31	0.45	0.27	2036	45.24	15.32	32.27	5.65	1.47	0.05
32	0.70	0.38	2036	6.83	70.09	10.41	11.10	1.52	0.05
33	0.52	0.19	2036	10.17	30.99	5.65	51.62	1.47	0.10
34	0.69	0.41	2036	10.27	12.52	69.30	6.39	1.47	0.05
35	0.48	0.28	2036	47.59	19.70	19.94	11.10	1.67	0.00

Table D.16
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Writing (CR)

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	4
1	2.44	0.86	0.77	2036	6.29	6.14	24.95	62.62	
2	2.31	0.92	0.79	2036	7.81	8.06	29.96	54.17	
3	2.29	0.94	0.80	2036	8.35	8.99	28.34	54.32	
4	2.07	0.99	0.74	2036	9.68	13.75	39.19	34.53	2.85

Table D.17
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Writing (MC)

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.50	0.32	2036	9.97	26.18	49.51	12.92	1.33	0.10
2	0.68	0.45	2036	10.66	5.65	68.47	13.26	1.92	0.05
3	0.52	0.43	2036	8.79	11.98	25.44	51.82	1.96	0.00
4	0.57	0.34	2036	56.97	14.00	9.63	17.04	2.36	0.00
5	0.68	0.40	2036	7.86	7.86	67.78	14.49	1.96	0.05
6	0.77	0.48	2036	6.19	76.92	7.37	7.51	2.01	0.00
7	0.77	0.63	2036	77.21	7.37	6.83	6.48	2.06	0.05
8	0.69	0.48	2036	68.91	14.19	5.21	9.63	2.01	0.05
9	0.57	0.43	2036	7.66	57.17	11.59	21.56	2.01	0.00
10	0.65	0.40	2036	8.55	10.66	64.54	14.00	2.21	0.05
11	0.78	0.58	2036	77.95	7.02	7.61	5.30	2.11	0.00
12	0.60	0.33	2036	18.66	12.23	59.58	7.37	2.16	0.00
13	0.72	0.50	2036	8.74	72.40	7.51	9.63	1.67	0.05
14	0.54	0.37	2036	19.35	10.90	13.65	53.83	2.26	0.00
15	0.33	0.21	2036	18.52	31.93	14.69	32.76	2.11	0.00

Table D.18a
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Speaking

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point			
					0	1	2	Omit
1	1.71	0.64	0.76	2036	6.53	8.69	81.04	3.73
2	1.70	0.64	0.85	2036	6.88	9.82	80.01	3.29
3	1.61	0.67	0.80	2036	7.56	18.12	71.51	2.80
4	1.23	0.82	0.65	2036	16.75	28.24	47.30	7.71
5	1.64	0.71	0.77	2036	7.51	8.69	77.60	6.19
6	1.76	0.57	0.84	2036	3.88	9.82	83.06	3.24
7	1.71	0.62	0.82	2036	5.80	10.71	80.11	3.39
8	1.45	0.75	0.75	2036	10.41	23.58	60.81	5.21
9	1.79	0.56	0.78	2036	3.93	6.34	86.15	3.59
10	1.76	0.57	0.82	2036	4.37	9.53	83.20	2.90
11	1.68	0.65	0.85	2036	6.43	12.38	77.65	3.54
12	1.62	0.68	0.83	2036	7.27	16.40	72.59	3.73

Table D.18b
2015 Operational Item Statistics
Cluster 4 Grades 6–8 Speaking Condition Codes

Item position	Percentage of respondents omitting item at each condition code					Total number of respondents omitting item
	A	B	C	D	No code	
1	5.26	38.16	21.06	13.16	22.36	76
2	1.50	35.82	11.94	17.92	32.84	67
3	3.50	36.84	7.02	12.28	40.36	57
4	4.46	33.76	29.30	17.84	14.64	157
5	2.38	42.86	26.20	9.52	19.04	126
6	9.10	28.78	9.10	15.16	37.88	66
7	5.80	37.68	7.24	13.04	36.24	69
8	7.54	42.46	14.16	13.20	22.64	106
9	2.74	35.62	10.96	17.80	32.88	73
10	6.78	30.50	3.38	16.94	42.38	59
11	2.78	31.94	15.28	15.28	34.72	72
12	6.58	35.52	13.16	10.52	34.22	76

Condition Codes:

- A** - inaudible/unclear
- B** - blank
- C** - refusal or excuse (“I don’t want to answer,” “I don’t know”)
- D** - off topic/off task/completely in language other than English

Table D.19
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Reading

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.42	0.53	2343	21.21	13.10	41.61	21.38	2.65	0.04
2	0.43	0.44	2343	43.19	19.80	20.83	13.36	2.82	0.00
3	0.48	0.46	2343	47.67	14.00	16.47	19.08	2.77	0.00
4	0.61	0.45	2343	10.16	20.87	61.29	5.08	2.60	0.00
5	0.37	0.30	2343	18.48	36.53	31.33	10.84	2.77	0.04
6	0.56	0.64	2343	9.56	21.43	56.00	10.29	2.73	0.00
7	0.63	0.45	2343	10.93	14.55	8.66	62.95	2.86	0.04
8	0.44	0.50	2343	26.89	9.35	44.17	16.77	2.73	0.09
9	0.41	0.51	2343	11.18	16.94	27.66	41.36	2.86	0.00
10	0.58	0.39	2343	58.47	24.24	5.29	8.88	3.07	0.04
11	0.37	0.44	2343	37.09	28.00	16.60	15.19	3.12	0.00
12	0.47	0.53	2343	25.48	47.03	12.51	12.25	2.73	0.00
13	0.43	0.44	2343	18.18	24.24	42.55	11.78	3.20	0.04
14	0.53	0.60	2343	14.98	53.48	16.22	12.51	2.82	0.00
15	0.63	0.51	2343	19.29	62.61	5.85	9.52	2.73	0.00
16	0.52	0.63	2343	14.81	21.94	52.24	8.19	2.77	0.04
17	0.43	0.51	2343	43.36	11.57	9.60	32.78	2.65	0.04
18	0.41	0.32	2343	40.93	15.92	15.58	24.71	2.82	0.04
19	0.48	0.49	2343	11.82	48.40	24.54	12.42	2.77	0.04
20	0.57	0.50	2343	13.10	13.96	12.89	57.19	2.77	0.09
21	0.56	0.51	2343	15.49	56.30	12.16	13.02	2.99	0.04
22	0.43	0.53	2343	19.50	43.45	6.40	27.49	3.07	0.09
23	0.49	0.54	2343	14.13	23.13	10.88	48.66	3.12	0.09
24	0.32	0.53	2343	9.09	35.34	20.91	31.71	2.94	0.00
25	0.43	0.52	2343	8.54	43.32	12.51	32.48	3.16	0.00
26	0.52	0.42	2343	13.96	23.47	51.52	7.85	3.16	0.04
27	0.35	0.49	2343	20.19	34.70	23.00	18.82	3.24	0.04
28	0.56	0.40	2343	15.28	55.95	13.83	11.10	3.84	0.00
29	0.58	0.52	2343	15.15	13.15	10.37	57.75	3.59	0.00
30	0.38	0.48	2343	21.51	14.94	22.62	37.52	3.41	0.00
31	0.53	0.36	2343	26.16	53.44	8.49	8.19	3.67	0.04
32	0.39	0.29	2343	16.90	39.48	18.27	21.17	4.18	0.00
33	0.40	0.43	2343	16.94	23.90	14.98	40.12	4.05	0.00
34	0.43	0.41	2343	12.76	11.95	27.57	43.36	4.31	0.04
35	0.50	0.30	2343	50.19	12.55	20.74	12.33	4.18	0.00

Table D.20
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Listening

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.52	0.59	2343	7.17	31.67	5.93	52.45	2.77	0.00
2	0.66	0.63	2343	6.15	12.89	66.41	11.65	2.90	0.00
3	0.63	0.45	2343	15.41	63.29	9.39	8.75	3.12	0.04
4	0.59	0.66	2343	58.73	14.04	14.04	10.12	2.94	0.13
5	0.50	0.64	2343	22.58	49.85	7.85	16.86	2.82	0.04
6	0.80	0.57	2343	79.81	8.24	4.91	4.27	2.73	0.04
7	0.60	0.42	2343	10.07	60.05	14.64	12.42	2.82	0.00
8	0.68	0.63	2343	12.46	13.19	67.56	3.88	2.82	0.09
9	0.61	0.54	2343	60.99	14.30	14.98	6.96	2.77	0.00
10	0.62	0.49	2343	8.15	15.45	11.18	62.31	2.82	0.09
11	0.57	0.66	2343	57.06	9.82	9.99	20.19	2.90	0.04
12	0.66	0.46	2343	19.50	6.32	65.73	5.63	2.77	0.04
13	0.54	0.47	2343	7.73	19.46	15.58	54.37	2.82	0.04
14	0.71	0.55	2343	71.15	3.80	7.21	15.07	2.69	0.09
15	0.51	0.60	2343	14.55	7.85	23.90	50.83	2.86	0.00
16	0.39	0.40	2343	32.05	15.88	39.01	9.99	2.99	0.09
17	0.41	0.38	2343	30.13	41.36	10.88	14.90	2.73	0.00
18	0.71	0.41	2343	6.79	70.81	6.23	13.49	2.69	0.00
19	0.63	0.48	2343	62.61	7.85	8.11	18.52	2.86	0.04
20	0.61	0.40	2343	11.69	61.42	7.00	17.20	2.65	0.04
21	0.63	0.56	2343	19.97	62.91	8.75	5.51	2.82	0.04
22	0.43	0.50	2343	13.32	42.98	17.11	23.73	2.77	0.09
23	0.42	0.30	2343	7.34	22.49	24.97	42.30	2.86	0.04
24	0.56	0.62	2343	15.19	16.35	9.90	55.87	2.69	0.00
25	0.50	0.58	2343	50.49	9.94	31.71	4.99	2.82	0.04
26	0.65	0.62	2343	20.49	65.17	8.11	3.41	2.82	0.00
27	0.38	0.49	2343	15.19	37.56	27.44	16.77	2.94	0.09
28	0.73	0.45	2343	6.36	5.34	12.42	72.98	2.90	0.00
29	0.47	0.36	2343	9.77	33.50	46.56	7.13	2.94	0.09
30	0.52	0.52	2343	23.52	52.03	15.15	6.23	3.07	0.00
31	0.68	0.52	2343	14.13	68.03	8.62	6.23	2.94	0.04
32	0.42	0.43	2343	27.53	9.13	41.83	18.40	3.07	0.04
33	0.36	0.29	2343	11.48	36.02	30.26	19.04	3.12	0.09
34	0.44	0.50	2343	12.76	21.38	44.34	18.22	3.24	0.04
35	0.68	0.56	2343	5.04	9.39	14.51	68.03	2.99	0.04

Table D.21
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Writing (CR)

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point				
					0	1	2	3	4
1	2.17	0.90	0.76	2343	8.54	8.19	41.02	42.25	
2	2.12	0.88	0.80	2343	7.43	11.14	43.70	37.73	
3	2.01	0.96	0.81	2343	11.57	10.76	42.68	35.00	
4	2.15	0.89	0.80	2343	8.11	9.09	42.34	40.46	
5	1.86	1.05	0.80	2343	14.98	14.72	42.51	24.80	2.99

Table D.22
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Writing (MC)

Item position	p-value	Point-biserial correlation	Number of respondents	Percentage of respondents at each score point					
				A	B	C	D	Omit	DB
1	0.62	0.57	2343	16.94	10.29	61.72	8.02	3.03	0.00
2	0.60	0.40	2343	60.35	9.35	15.02	12.21	3.07	0.00
3	0.42	0.49	2343	27.83	22.62	3.97	42.42	3.12	0.04
4	0.51	0.45	2343	20.32	11.91	50.83	13.10	3.80	0.04
5	0.48	0.44	2343	20.19	19.12	47.67	9.94	3.07	0.00
6	0.49	0.44	2343	18.35	48.78	12.97	16.65	3.20	0.04
7	0.43	0.22	2343	17.88	19.72	43.28	15.83	3.29	0.00
8	0.47	0.37	2343	21.47	47.25	16.35	11.61	3.33	0.00
9	0.42	0.39	2343	9.73	28.68	16.77	41.57	3.20	0.04
10	0.37	0.33	2343	37.30	21.68	10.37	27.23	3.41	0.00
11	0.38	0.34	2343	18.69	37.60	22.19	18.31	3.20	0.00
12	0.58	0.47	2343	9.09	57.87	15.49	14.08	3.46	0.00
13	0.39	0.29	2343	25.35	39.10	24.54	7.68	3.33	0.00
14	0.55	0.46	2343	14.98	14.77	11.87	54.89	3.46	0.04
15	0.51	0.42	2343	21.17	12.21	12.29	50.83	3.46	0.04

Table D.23a
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Speaking

Item position	Mean score	SD	Item-total correlation	Number of respondents	Percentage of respondents at each score point			
					0	1	2	Omit
1	1.48	0.84	0.82	2343	10.37	6.15	71.02	12.46
2	1.58	0.74	0.81	2343	8.58	12.42	72.56	6.44
3	1.31	0.83	0.82	2343	15.58	22.15	54.59	7.68
4	1.27	0.88	0.84	2343	14.85	14.47	56.47	14.21
5	1.60	0.76	0.78	2343	7.90	6.91	76.31	8.88
6	1.64	0.68	0.78	2343	5.42	12.72	75.42	6.44
7	1.49	0.77	0.82	2343	9.69	17.41	65.73	7.17
8	1.11	0.86	0.81	2343	18.69	23.94	43.75	13.62
9	1.23	0.91	0.76	2343	17.58	12.63	55.01	14.77
10	1.45	0.77	0.76	2343	10.67	21.43	61.59	6.32
11	1.27	0.83	0.81	2343	17.29	23.77	51.69	7.26
12	1.11	0.86	0.81	2343	19.59	24.93	43.02	12.42

Table D.23b
2015 Operational Item Statistics
Cluster 5 Grades 9–12 Speaking Condition Codes

Item position	Percentage of respondents omitting item at each condition code					Total number of respondents omitting item
	A	B	C	D	No code	
1	4.80	36.98	27.06	7.88	23.28	292
2	7.28	25.82	13.90	7.94	45.04	151
3	6.12	28.34	15.56	8.34	41.66	180
4	2.70	32.74	31.24	11.12	22.22	333
5	2.88	36.54	16.34	8.66	35.58	208
6	1.98	29.14	11.92	7.28	49.66	151
7	4.16	27.38	13.70	10.12	44.64	168
8	3.44	34.16	30.40	8.16	23.82	319
9	4.92	35.26	25.72	12.14	21.96	346
10	6.08	28.38	7.44	7.44	50.68	148
11	5.88	29.42	11.18	9.42	44.12	170
12	3.44	29.56	24.74	14.44	27.84	291

Condition Codes:

- A** - inaudible/unclear
- B** - blank
- C** - refusal or excuse (“I don’t want to answer,” “I don’t know”)
- D** - off topic/off task/completely in language other than English

Appendix E

2015 Raw Score Frequency Tables

Table E.1
2015 Operational Raw Score Frequencies
Kindergarten Reading

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	24	0.72	24	0.72
1	20	0.60	44	1.32
2	35	1.05	79	2.37
3	16	0.48	95	2.85
4	26	0.78	121	3.63
5	23	0.69	144	4.32
6	23	0.69	167	5.01
7	28	0.84	195	5.85
8	35	1.05	230	6.90
9	38	1.14	268	8.04
10	51	1.53	319	9.57
11	62	1.86	381	11.43
12	60	1.80	441	13.23
13	60	1.80	501	15.03
14	76	2.28	577	17.31
15	61	1.83	638	19.14
16	80	2.40	718	21.54
17	75	2.25	793	23.79
18	112	3.36	905	27.14
19	81	2.43	986	29.57
20	118	3.54	1104	33.11
21	122	3.66	1226	36.77
22	117	3.51	1343	40.28
23	105	3.15	1448	43.43
24	103	3.09	1551	46.52
25	128	3.84	1679	50.36
26	102	3.06	1781	53.42
27	125	3.75	1906	57.17
28	114	3.42	2020	60.59
29	116	3.48	2136	64.07
30	106	3.18	2242	67.25
31	143	4.29	2385	71.54
32	94	2.82	2479	74.36
33	129	3.87	2608	78.22
34	108	3.24	2716	81.46
35	95	2.85	2811	84.31
36	77	2.31	2888	86.62
37	106	3.18	2994	89.80
38	77	2.31	3071	92.11
39	96	2.88	3167	94.99
40	51	1.53	3218	96.52
41	63	1.89	3281	98.41
42	53	1.59	3334	100.00

Table E.2
2015 Operational Raw Score Frequencies
Kindergarten Listening

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	40	1.20	40	1.20
1	43	1.29	83	2.49
2	27	0.81	110	3.30
3	42	1.26	152	4.56
4	101	3.03	253	7.59
5	80	2.40	333	9.99
6	90	2.70	423	12.69
7	226	6.78	649	19.47
8	186	5.58	835	25.04
9	173	5.19	1008	30.23
10	198	5.94	1206	36.17
11	220	6.60	1426	42.77
12	191	5.73	1617	48.50
13	244	7.32	1861	55.82
14	288	8.64	2149	64.46
15	203	6.09	2352	70.55
16	217	6.51	2569	77.05
17	186	5.58	2755	82.63
18	166	4.98	2921	87.61
19	125	3.75	3046	91.36
20	82	2.46	3128	93.82
21	206	6.18	3334	100.00

Table E.3
2015 Operational Raw Score Frequencies
Kindergarten Writing

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	40	1.20	40	1.20
1	35	1.05	75	2.25
2	64	1.92	139	4.17
3	47	1.41	186	5.58
4	73	2.19	259	7.77
5	81	2.43	340	10.20
6	88	2.64	428	12.84
7	105	3.15	533	15.99
8	110	3.30	643	19.29
9	155	4.65	798	23.94
10	166	4.98	964	28.91
11	147	4.41	1111	33.32
12	177	5.31	1288	38.63
13	153	4.59	1441	43.22
14	150	4.50	1591	47.72
15	165	4.95	1756	52.67
16	183	5.49	1939	58.16
17	192	5.76	2131	63.92
18	214	6.42	2345	70.34
19	195	5.85	2540	76.18
20	143	4.29	2683	80.47
21	154	4.62	2837	85.09
22	122	3.66	2959	88.75
23	99	2.97	3058	91.72
24	97	2.91	3155	94.63
25	66	1.98	3221	96.61
26	58	1.74	3279	98.35
27	55	1.65	3334	100.00

Table E.4
2015 Operational Raw Score Frequencies
Kindergarten Speaking

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	51	1.53	51	1.53
1	19	0.57	70	2.10
2	30	0.90	100	3.00
3	27	0.81	127	3.81
4	24	0.72	151	4.53
5	56	1.68	207	6.21
6	55	1.65	262	7.86
7	62	1.86	324	9.72
8	179	5.37	503	15.09
9	115	3.45	618	18.54
10	109	3.27	727	21.81
11	132	3.96	859	25.76
12	113	3.39	972	29.15
13	156	4.68	1128	33.83
14	185	5.55	1313	39.38
15	205	6.15	1518	45.53
16	362	10.86	1880	56.39
17	214	6.42	2094	62.81
18	214	6.42	2308	69.23
19	188	5.64	2496	74.87
20	190	5.70	2686	80.56
21	167	5.01	2853	85.57
22	123	3.69	2976	89.26
23	108	3.24	3084	92.50
24	250	7.50	3334	100.00

Table E.5
2015 Operational Raw Score Frequencies
Grades 1–2 Reading

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	39	0.70	39	0.70
1	24	0.43	63	1.14
2	21	0.38	84	1.52
3	17	0.31	101	1.82
4	16	0.29	117	2.11
5	37	0.67	154	2.78
6	35	0.63	189	3.41
7	28	0.51	217	3.92
8	44	0.79	261	4.71
9	40	0.72	301	5.43
10	50	0.90	351	6.33
11	44	0.79	395	7.13
12	49	0.88	444	8.01
13	41	0.74	485	8.75
14	81	1.46	566	10.21
15	90	1.62	656	11.84
16	101	1.82	757	13.66
17	87	1.57	844	15.23
18	131	2.36	975	17.60
19	144	2.60	1119	20.19
20	151	2.73	1270	22.92
21	156	2.82	1426	25.74
22	150	2.71	1576	28.44
23	171	3.09	1747	31.53
24	180	3.25	1927	34.78
25	166	3.00	2093	37.77
26	162	2.92	2255	40.70
27	181	3.27	2436	43.96
28	232	4.19	2668	48.15
29	233	4.21	2901	52.36
30	225	4.06	3126	56.42
31	225	4.06	3351	60.48
32	252	4.55	3603	65.02
33	247	4.46	3850	69.48
34	238	4.30	4088	73.78
35	199	3.59	4287	77.37
36	188	3.39	4475	80.76
37	177	3.19	4652	83.96
38	185	3.34	4837	87.29
39	171	3.09	5008	90.38
40	147	2.65	5155	93.03
41	135	2.44	5290	95.47
42	251	4.53	5541	100.00

Table E.6
2015 Operational Raw Score Frequencies
Grades 1–2 Listening

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	46	0.83	46	0.83
1	19	0.34	65	1.17
2	25	0.45	90	1.62
3	43	0.78	133	2.40
4	64	1.16	197	3.56
5	64	1.16	261	4.71
6	66	1.19	327	5.90
7	165	2.98	492	8.88
8	176	3.18	668	12.06
9	211	3.81	879	15.86
10	222	4.01	1101	19.87
11	305	5.50	1406	25.37
12	269	4.85	1675	30.23
13	313	5.65	1988	35.88
14	529	9.55	2517	45.43
15	498	8.99	3015	54.41
16	482	8.70	3497	63.11
17	421	7.60	3918	70.71
18	380	6.86	4298	77.57
19	314	5.67	4612	83.23
20	274	4.94	4886	88.18
21	655	11.82	5541	100.00

Table E.7
2015 Operational Raw Score Frequencies
Grades 1–2 Writing

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	44	0.79	44	0.79
1	16	0.29	60	1.08
2	23	0.42	83	1.50
3	25	0.45	108	1.95
4	49	0.88	157	2.83
5	49	0.88	206	3.72
6	39	0.70	245	4.42
7	48	0.87	293	5.29
8	68	1.23	361	6.52
9	101	1.82	462	8.34
10	100	1.80	562	10.14
11	122	2.20	684	12.34
12	157	2.83	841	15.18
13	164	2.96	1005	18.14
14	192	3.47	1197	21.60
15	203	3.66	1400	25.27
16	202	3.65	1602	28.91
17	294	5.31	1896	34.22
18	393	7.09	2289	41.31
19	423	7.63	2712	48.94
20	479	8.64	3191	57.59
21	475	8.57	3666	66.16
22	469	8.46	4135	74.63
23	427	7.71	4562	82.33
24	347	6.26	4909	88.59
25	271	4.89	5180	93.48
26	166	3.00	5346	96.48
27	195	3.52	5541	100.00

Table E.8
2015 Operational Raw Score Frequencies
Grades 1–2 Speaking

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	59	1.06	59	1.06
1	21	0.38	80	1.44
2	15	0.27	95	1.71
3	23	0.42	118	2.13
4	16	0.29	134	2.42
5	43	0.78	177	3.19
6	57	1.03	234	4.22
7	71	1.28	305	5.50
8	141	2.54	446	8.05
9	90	1.62	536	9.67
10	93	1.68	629	11.35
11	117	2.11	746	13.46
12	136	2.45	882	15.92
13	164	2.96	1046	18.88
14	214	3.86	1260	22.74
15	249	4.49	1509	27.23
16	522	9.42	2031	36.65
17	424	7.65	2455	44.31
18	444	8.01	2899	52.32
19	473	8.54	3372	60.86
20	410	7.40	3782	68.25
21	396	7.15	4178	75.40
22	436	7.87	4614	83.27
23	357	6.44	4971	89.71
24	570	10.29	5541	100.00

Table E.9
2015 Operational Raw Score Frequencies
Grades 3–5 Reading

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	17	0.32	17	0.32
1	12	0.23	29	0.55
2	8	0.15	37	0.70
3	11	0.21	48	0.91
4	3	0.06	51	0.97
5	18	0.34	69	1.31
6	19	0.36	88	1.67
7	31	0.59	119	2.26
8	40	0.76	159	3.03
9	61	1.16	220	4.19
10	73	1.39	293	5.58
11	66	1.26	359	6.83
12	86	1.64	445	8.47
13	80	1.52	525	9.99
14	103	1.96	628	11.95
15	78	1.48	706	13.44
16	94	1.79	800	15.23
17	108	2.06	908	17.28
18	118	2.25	1026	19.53
19	151	2.87	1177	22.40
20	143	2.72	1320	25.12
21	177	3.37	1497	28.49
22	167	3.18	1664	31.67
23	197	3.75	1861	35.42
24	207	3.94	2068	39.36
25	285	5.42	2353	44.78
26	287	5.46	2640	50.25
27	302	5.75	2942	56.00
28	298	5.67	3240	61.67
29	317	6.03	3557	67.70
30	373	7.10	3930	74.80
31	357	6.79	4287	81.59
32	374	7.12	4661	88.71
33	288	5.48	4949	94.19
34	210	4.00	5159	98.19
35	95	1.81	5254	100.00

Table E.10
2015 Operational Raw Score Frequencies
Grades 3–5 Listening

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	19	0.36	19	0.36
1	5	0.10	24	0.46
2	13	0.25	37	0.70
3	10	0.19	47	0.89
4	4	0.08	51	0.97
5	13	0.25	64	1.22
6	25	0.48	89	1.69
7	22	0.42	111	2.11
8	28	0.53	139	2.65
9	44	0.84	183	3.48
10	39	0.74	222	4.23
11	44	0.84	266	5.06
12	50	0.95	316	6.01
13	63	1.20	379	7.21
14	89	1.69	468	8.91
15	88	1.67	556	10.58
16	114	2.17	670	12.75
17	117	2.23	787	14.98
18	143	2.72	930	17.70
19	186	3.54	1116	21.24
20	189	3.60	1305	24.84
21	207	3.94	1512	28.78
22	273	5.20	1785	33.97
23	305	5.81	2090	39.78
24	350	6.66	2440	46.44
25	359	6.83	2799	53.27
26	432	8.22	3231	61.50
27	392	7.46	3623	68.96
28	385	7.33	4008	76.28
29	342	6.51	4350	82.79
30	323	6.15	4673	88.94
31	262	4.99	4935	93.93
32	162	3.08	5097	97.01
33	104	1.98	5201	98.99
34	39	0.74	5240	99.73
35	14	0.27	5254	100.00

Table E.11
2015 Operational Raw Score Frequencies
Grades 3–5 Writing

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	31	0.59	31	0.59
1	10	0.19	41	0.78
2	18	0.34	59	1.12
3	32	0.61	91	1.73
4	34	0.65	125	2.38
5	33	0.63	158	3.01
6	36	0.69	194	3.69
7	44	0.84	238	4.53
8	61	1.16	299	5.69
9	94	1.79	393	7.48
10	83	1.58	476	9.06
11	114	2.17	590	11.23
12	174	3.31	764	14.54
13	208	3.96	972	18.50
14	259	4.93	1231	23.43
15	278	5.29	1509	28.72
16	352	6.70	1861	35.42
17	347	6.60	2208	42.03
18	383	7.29	2591	49.31
19	420	7.99	3011	57.31
20	448	8.53	3459	65.84
21	433	8.24	3892	74.08
22	406	7.73	4298	81.80
23	365	6.95	4663	88.75
24	261	4.97	4924	93.72
25	180	3.43	5104	97.15
26	102	1.94	5206	99.09
27	41	0.78	5247	99.87
28	7	0.13	5254	100.00

Table E.12
2015 Operational Raw Score Frequencies
Grades 3–5 Speaking

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	136	2.59	136	2.59
1	20	0.38	156	2.97
2	8	0.15	164	3.12
3	19	0.36	183	3.48
4	15	0.29	198	3.77
5	9	0.17	207	3.94
6	16	0.30	223	4.24
7	15	0.29	238	4.53
8	17	0.32	255	4.85
9	23	0.44	278	5.29
10	21	0.40	299	5.69
11	29	0.55	328	6.24
12	41	0.78	369	7.02
13	50	0.95	419	7.97
14	82	1.56	501	9.54
15	83	1.58	584	11.12
16	110	2.09	694	13.21
17	161	3.06	855	16.27
18	244	4.64	1099	20.92
19	346	6.59	1445	27.50
20	523	9.95	1968	37.46
21	713	13.57	2681	51.03
22	848	16.14	3529	67.17
23	876	16.67	4405	83.84
24	849	16.16	5254	100.00

Table E.13
2015 Operational Raw Score Frequencies
Grades 6–8 Reading

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	15	0.74	15	0.74
1	13	0.64	28	1.38
2	2	0.10	30	1.47
3	2	0.10	32	1.57
4	7	0.34	39	1.92
5	7	0.34	46	2.26
6	12	0.59	58	2.85
7	25	1.23	83	4.08
8	16	0.79	99	4.86
9	40	1.96	139	6.83
10	51	2.50	190	9.33
11	49	2.41	239	11.74
12	56	2.75	295	14.49
13	56	2.75	351	17.24
14	49	2.41	400	19.65
15	53	2.60	453	22.25
16	47	2.31	500	24.56
17	42	2.06	542	26.62
18	56	2.75	598	29.37
19	52	2.55	650	31.93
20	44	2.16	694	34.09
21	59	2.90	753	36.98
22	52	2.55	805	39.54
23	57	2.80	862	42.34
24	73	3.59	935	45.92
25	64	3.14	999	49.07
26	86	4.22	1085	53.29
27	103	5.06	1188	58.35
28	109	5.35	1297	63.70
29	138	6.78	1435	70.48
30	135	6.63	1570	77.11
31	134	6.58	1704	83.69
32	131	6.43	1835	90.13
33	114	5.60	1949	95.73
34	54	2.65	2003	98.38
35	33	1.62	2036	100.00

Table E.14
2015 Operational Raw Score Frequencies
Grades 6–8 Listening

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	18	0.88	18	0.88
1	7	0.34	25	1.23
2	1	0.05	26	1.28
3	3	0.15	29	1.42
4	2	0.10	31	1.52
5	5	0.25	36	1.77
6	23	1.13	59	2.90
7	18	0.88	77	3.78
8	25	1.23	102	5.01
9	27	1.33	129	6.34
10	35	1.72	164	8.06
11	22	1.08	186	9.14
12	33	1.62	219	10.76
13	26	1.28	245	12.03
14	20	0.98	265	13.02
15	21	1.03	286	14.05
16	26	1.28	312	15.32
17	27	1.33	339	16.65
18	37	1.82	376	18.47
19	42	2.06	418	20.53
20	65	3.19	483	23.72
21	40	1.96	523	25.69
22	57	2.80	580	28.49
23	72	3.54	652	32.02
24	90	4.42	742	36.44
25	85	4.17	827	40.62
26	104	5.11	931	45.73
27	134	6.58	1065	52.31
28	147	7.22	1212	59.53
29	137	6.73	1349	66.26
30	190	9.33	1539	75.59
31	181	8.89	1720	84.48
32	134	6.58	1854	91.06
33	111	5.45	1965	96.51
34	49	2.41	2014	98.92
35	22	1.08	2036	100.00

Table E.15
2015 Operational Raw Score Frequencies
Grades 6–8 Writing

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	26	1.28	26	1.28
1	5	0.25	31	1.52
2	19	0.93	50	2.46
3	14	0.69	64	3.14
4	26	1.28	90	4.42
5	27	1.33	117	5.75
6	32	1.57	149	7.32
7	36	1.77	185	9.09
8	30	1.47	215	10.56
9	37	1.82	252	12.38
10	33	1.62	285	14.00
11	42	2.06	327	16.06
12	51	2.50	378	18.57
13	62	3.05	440	21.61
14	44	2.16	484	23.77
15	50	2.46	534	26.23
16	69	3.39	603	29.62
17	69	3.39	672	33.01
18	104	5.11	776	38.11
19	133	6.53	909	44.65
20	132	6.48	1041	51.13
21	127	6.24	1168	57.37
22	211	10.36	1379	67.73
23	196	9.63	1575	77.36
24	176	8.64	1751	86.00
25	146	7.17	1897	93.17
26	98	4.81	1995	97.99
27	36	1.77	2031	99.75
28	5	0.25	2036	100.00

Table E.16
2015 Operational Raw Score Frequencies
Grades 6–8 Speaking

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	85	4.17	85	4.17
1	16	0.79	101	4.96
2	17	0.83	118	5.80
3	15	0.74	133	6.53
4	21	1.03	154	7.56
5	12	0.59	166	8.15
6	9	0.44	175	8.60
7	12	0.59	187	9.18
8	12	0.59	199	9.77
9	20	0.98	219	10.76
10	9	0.44	228	11.20
11	16	0.79	244	11.98
12	18	0.88	262	12.87
13	18	0.88	280	13.75
14	26	1.28	306	15.03
15	27	1.33	333	16.36
16	29	1.42	362	17.78
17	32	1.57	394	19.35
18	62	3.05	456	22.40
19	77	3.78	533	26.18
20	109	5.35	642	31.53
21	160	7.86	802	39.39
22	259	12.72	1061	52.11
23	368	18.07	1429	70.19
24	607	29.81	2036	100.00

Table E.17
2015 Operational Raw Score Frequencies
Grades 9–12 Reading

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	60	2.56	60	2.56
1	0	0.00	60	2.56
2	3	0.13	63	2.69
3	5	0.21	68	2.90
4	32	1.37	100	4.27
5	44	1.88	144	6.15
6	94	4.01	238	10.16
7	90	3.84	328	14.00
8	135	5.76	463	19.76
9	141	6.02	604	25.78
10	130	5.55	734	31.33
11	128	5.46	862	36.79
12	94	4.01	956	40.80
13	94	4.01	1050	44.81
14	79	3.37	1129	48.19
15	73	3.12	1202	51.30
16	70	2.99	1272	54.29
17	76	3.24	1348	57.53
18	62	2.65	1410	60.18
19	51	2.18	1461	62.36
20	68	2.90	1529	65.26
21	56	2.39	1585	67.65
22	63	2.69	1648	70.34
23	65	2.77	1713	73.11
24	62	2.65	1775	75.76
25	72	3.07	1847	78.83
26	69	2.94	1916	81.78
27	66	2.82	1982	84.59
28	62	2.65	2044	87.24
29	66	2.82	2110	90.06
30	52	2.22	2162	92.27
31	47	2.01	2209	94.28
32	60	2.56	2269	96.84
33	38	1.62	2307	98.46
34	25	1.07	2332	99.53
35	11	0.47	2343	100.00

Table E.18
2015 Operational Raw Score Frequencies
Grades 9–12 Listening

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	62	2.65	62	2.65
1	0	0.00	62	2.65
2	2	0.09	64	2.73
3	9	0.38	73	3.12
4	13	0.55	86	3.67
5	25	1.07	111	4.74
6	44	1.88	155	6.62
7	74	3.16	229	9.77
8	83	3.54	312	13.32
9	75	3.20	387	16.52
10	83	3.54	470	20.06
11	79	3.37	549	23.43
12	71	3.03	620	26.46
13	79	3.37	699	29.83
14	68	2.90	767	32.74
15	73	3.12	840	35.85
16	73	3.12	913	38.97
17	70	2.99	983	41.95
18	63	2.69	1046	44.64
19	55	2.35	1101	46.99
20	60	2.56	1161	49.55
21	77	3.29	1238	52.84
22	74	3.16	1312	56.00
23	86	3.67	1398	59.67
24	67	2.86	1465	62.53
25	86	3.67	1551	66.20
26	72	3.07	1623	69.27
27	93	3.97	1716	73.24
28	94	4.01	1810	77.25
29	101	4.31	1911	81.56
30	103	4.40	2014	85.96
31	76	3.24	2090	89.20
32	87	3.71	2177	92.92
33	64	2.73	2241	95.65
34	68	2.90	2309	98.55
35	34	1.45	2343	100.00

Table E.19
2015 Operational Raw Score Frequencies
Grades 9–12 Writing

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	62	2.65	62	2.65
1	3	0.13	65	2.77
2	12	0.51	77	3.29
3	19	0.81	96	4.10
4	28	1.20	124	5.29
5	34	1.45	158	6.74
6	39	1.66	197	8.41
7	44	1.88	241	10.29
8	48	2.05	289	12.33
9	65	2.77	354	15.11
10	68	2.90	422	18.01
11	67	2.86	489	20.87
12	91	3.88	580	24.75
13	94	4.01	674	28.77
14	116	4.95	790	33.72
15	108	4.61	898	38.33
16	115	4.91	1013	43.24
17	107	4.57	1120	47.80
18	111	4.74	1231	52.54
19	100	4.27	1331	56.81
20	108	4.61	1439	61.42
21	113	4.82	1552	66.24
22	102	4.35	1654	70.59
23	100	4.27	1754	74.86
24	99	4.23	1853	79.09
25	119	5.08	1972	84.17
26	104	4.44	2076	88.60
27	94	4.01	2170	92.62
28	84	3.59	2254	96.20
29	51	2.18	2305	98.38
30	26	1.11	2331	99.49
31	12	0.51	2343	100.00

Table E.20
2015 Operational Raw Score Frequencies
Grades 9–12 Speaking

Raw Score	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
0	191	8.15	191	8.15
1	37	1.58	228	9.73
2	31	1.32	259	11.05
3	39	1.66	298	12.72
4	31	1.32	329	14.04
5	39	1.66	368	15.71
6	50	2.13	418	17.84
7	33	1.41	451	19.25
8	45	1.92	496	21.17
9	41	1.75	537	22.92
10	42	1.79	579	24.71
11	39	1.66	618	26.38
12	44	1.88	662	28.25
13	41	1.75	703	30.00
14	51	2.18	754	32.18
15	42	1.79	796	33.97
16	67	2.86	863	36.83
17	66	2.82	929	39.65
18	80	3.41	1009	43.06
19	78	3.33	1087	46.39
20	133	5.68	1220	52.07
21	136	5.80	1356	57.87
22	196	8.37	1552	66.24
23	242	10.33	1794	76.57
24	549	23.43	2343	100.00

Appendix F

2015 Operational Test Summary Statistics

Table F.1
2015 Operational Test Summary Statistics

Grade Cluster	Test	Number of respondents	Maximum Score	Raw Score Mean	Raw Score SD	Reliability (α)	SEM
K	Reading	3334	42	24.70	10.04	0.96	2.22
	Listening	3334	21	12.32	5.10	0.94	1.18
	Writing	3334	27	14.48	6.48	0.94	1.54
	Speaking	3334	24	15.20	5.76	0.96	1.24
1–2	Reading	5541	42	27.54	9.66	0.96	1.88
	Listening	5541	21	14.50	4.74	0.94	1.12
	Writing	5541	27	18.48	5.68	0.94	1.44
	Speaking	5541	24	17.40	5.24	0.94	1.22
3–5	Reading	5254	35	24.76	7.36	0.90	2.22
	Listening	5254	35	23.82	6.26	0.88	2.26
	Writing	5254	28	17.80	5.20	0.80	2.32
	Speaking	5254	24	19.96	5.00	0.92	1.40
6–8	Reading	2036	35	23.26	8.38	0.92	2.30
	Listening	2036	35	24.78	7.68	0.92	2.22
	Writing	2036	28	18.46	6.44	0.88	2.26
	Speaking	2036	24	19.64	6.52	0.96	1.32
9–12	Reading	2343	35	16.66	8.80	0.92	2.54
	Listening	2343	35	19.82	9.20	0.94	2.42
	Writing	2343	31	17.52	7.34	0.88	2.52
	Speaking	2343	24	16.54	8.14	0.96	1.60

Appendix G

2015 Operational Raw Score Summary Statistics

Table G.1
2015 Operational Raw Score Summary Statistics – Grade Cluster K

Domain	Group	Raw Score Summary Statistics		
		N	Mean	STD
Reading	Total	3334	24.70	10.04
	Male	1753	23.93	10.06
	Female	1581	25.55	9.95
	Asian	355	24.03	11.42
	Black	223	27.23	9.57
	Hispanic	2521	24.30	9.77
	White	196	28.07	10.22
	American Indian/Alaska Native	NR	NR	NR
	Native Hawaiian/Pacific Islander	10	24.10	10.20
Listening	Total	3334	12.33	5.11
	Male	1753	11.94	5.14
	Female	1581	12.75	5.04
	Asian	355	11.55	5.80
	Black	223	13.57	5.25
	Hispanic	2521	12.25	4.93
	White	196	13.46	5.41
	American Indian/Alaska Native	NR	NR	NR
	Native Hawaiian/Pacific Islander	10	10.90	5.02
Writing	Total	3334	14.48	6.48
	Male	1753	13.93	6.42
	Female	1581	15.10	6.49
	Asian	355	14.46	7.48
	Black	223	16.08	6.33
	Hispanic	2521	14.12	6.27
	White	196	17.12	6.51
	American Indian/Alaska Native	NR	NR	NR
	Native Hawaiian/Pacific Islander	10	13.50	6.00
Speaking	Total	3334	15.21	5.76
	Male	1753	14.89	5.81
	Female	1581	15.56	5.68
	Asian	355	13.94	6.82
	Black	223	16.83	5.52
	Hispanic	2521	15.19	5.55
	White	196	16.32	5.77
	American Indian/Alaska Native	NR	NR	NR
	Native Hawaiian/Pacific Islander	10	9.70	6.09

Table G.2
2015 Operational Raw Score Summary Statistics – Grade Cluster 1–2

Domain	Group	Raw Score Summary Statistics		
		N	Mean	STD
Reading	Total	5541	27.54	9.65
	Male	2892	27.10	9.89
	Female	2649	28.02	9.37
	Asian	623	26.89	11.47
	Black	344	29.03	9.86
	Hispanic	4165	27.44	9.26
	White	329	28.40	10.23
	American Indian/Alaska Native	24	30.83	8.51
	Native Hawaiian/Pacific Islander	14	23.64	10.59
Listening	Total	5541	14.49	4.75
	Male	2892	14.24	4.85
	Female	2649	14.76	4.62
	Asian	623	13.97	5.50
	Black	344	15.01	4.84
	Hispanic	4165	14.49	4.60
	White	329	14.79	4.85
	American Indian/Alaska Native	24	15.96	3.86
	Native Hawaiian/Pacific Islander	14	12.14	5.84
Writing	Total	5541	18.48	5.68
	Male	2892	17.91	5.77
	Female	2649	19.11	5.50
	Asian	623	18.20	6.61
	Black	344	18.61	5.88
	Hispanic	4165	18.46	5.48
	White	329	19.20	5.89
	American Indian/Alaska Native	24	20.63	3.72
	Native Hawaiian/Pacific Islander	14	15.86	6.98
Speaking	Total	5541	17.40	5.23
	Male	2892	17.13	5.40
	Female	2649	17.70	5.03
	Asian	623	15.83	6.22
	Black	344	18.32	5.07
	Hispanic	4165	17.53	5.01
	White	329	17.79	5.61
	American Indian/Alaska Native	24	19.38	3.77
	Native Hawaiian/Pacific Islander	14	15.07	5.53

Table G.3
2015 Operational Raw Score Summary Statistics – Grade Cluster 3–5

Domain	Group	Raw Score Summary Statistics		
		N	Mean	STD
Reading	Total	5254	24.75	7.36
	Male	2725	24.30	7.60
	Female	2529	25.25	7.06
	Asian	593	23.58	8.88
	Black	319	24.26	7.37
	Hispanic	3943	25.09	6.97
	White	333	23.20	8.33
	American Indian/Alaska Native	35	24.97	8.87
	Native Hawaiian/Pacific Islander	11	28.09	4.91
Listening	Total	5254	23.82	6.26
	Male	2725	23.70	6.44
	Female	2529	23.94	6.07
	Asian	593	22.06	7.10
	Black	319	23.56	5.98
	Hispanic	3943	24.20	6.01
	White	333	22.67	7.00
	American Indian/Alaska Native	35	24.20	8.54
	Native Hawaiian/Pacific Islander	11	24.09	4.50
Writing	Total	5254	17.79	5.20
	Male	2725	17.20	5.24
	Female	2529	18.43	5.09
	Asian	593	16.20	5.68
	Black	319	17.56	5.13
	Hispanic	3943	18.10	4.95
	White	333	17.07	6.44
	American Indian/Alaska Native	35	18.34	6.66
	Native Hawaiian/Pacific Islander	11	19.45	3.91
Speaking	Total	5254	19.96	5.00
	Male	2725	19.85	5.00
	Female	2529	20.08	4.99
	Asian	593	18.87	5.88
	Black	319	20.31	4.10
	Hispanic	3943	20.16	4.82
	White	333	19.28	5.69
	American Indian/Alaska Native	35	19.46	6.39
	Native Hawaiian/Pacific Islander	11	20.73	3.80

Table G.4
2015 Operational Raw Score Summary Statistics – Grade Cluster 6–8

Domain	Group	Raw Score Summary Statistics		
		N	Mean	STD
Reading	Total	2036	23.26	8.38
	Male	1112	22.58	8.66
	Female	924	24.08	7.96
	Asian	319	20.16	8.77
	Black	154	24.12	7.31
	Hispanic	1376	24.12	8.07
	White	137	20.66	9.38
	American Indian/Alaska Native	27	23.89	9.29
	Native Hawaiian/Pacific Islander	11	25.18	5.38
Listening	Total	2036	24.78	7.68
	Male	1112	24.29	8.02
	Female	924	25.39	7.21
	Asian	319	22.19	8.44
	Black	154	26.28	5.91
	Hispanic	1376	25.34	7.53
	White	137	23.50	7.73
	American Indian/Alaska Native	27	24.30	8.57
	Native Hawaiian/Pacific Islander	11	27.36	4.52
Writing	Total	2036	18.46	6.44
	Male	1112	17.70	6.59
	Female	924	19.38	6.13
	Asian	319	16.07	6.57
	Black	154	19.73	5.18
	Hispanic	1376	18.98	6.27
	White	137	17.15	7.58
	American Indian/Alaska Native	27	19.33	6.83
	Native Hawaiian/Pacific Islander	11	19.09	5.07
Speaking	Total	2036	19.64	6.52
	Male	1112	19.26	6.93
	Female	924	20.11	5.97
	Asian	319	17.69	6.97
	Black	154	21.38	4.27
	Hispanic	1376	20.02	6.48
	White	137	18.68	6.84
	American Indian/Alaska Native	27	17.89	7.82
	Native Hawaiian/Pacific Islander	11	21.36	2.46

Table G.5
2015 Operational Raw Score Summary Statistics – Grade Cluster 9–12

Domain	Group	Raw Score Summary Statistics		
		N	Mean	STD
Reading	Total	2343	16.66	8.80
	Male	1318	16.37	8.91
	Female	1025	17.03	8.64
	Asian	519	15.13	7.53
	Black	177	19.19	8.61
	Hispanic	1420	16.94	9.23
	White	180	16.10	8.06
	American Indian/Alaska Native	13	16.08	11.76
	Native Hawaiian/Pacific Islander	23	17.87	8.16
Listening	Total	2343	19.83	9.20
	Male	1318	19.39	9.44
	Female	1025	20.40	8.85
	Asian	519	18.97	8.33
	Black	177	22.93	7.97
	Hispanic	1420	19.60	9.64
	White	180	20.93	8.36
	American Indian/Alaska Native	13	18.77	11.78
	Native Hawaiian/Pacific Islander	23	21.52	8.65
Writing	Total	2343	17.53	7.35
	Male	1318	17.07	7.57
	Female	1025	18.11	7.01
	Asian	519	17.09	6.31
	Black	177	20.28	6.42
	Hispanic	1420	17.25	7.70
	White	180	18.00	7.59
	American Indian/Alaska Native	13	15.92	9.27
	Native Hawaiian/Pacific Islander	23	20.30	5.49
Speaking	Total	2343	16.53	8.15
	Male	1318	16.27	8.30
	Female	1025	16.87	7.94
	Asian	519	16.83	7.22
	Black	177	20.29	5.81
	Hispanic	1420	15.88	8.67
	White	180	17.17	7.41
	American Indian/Alaska Native	13	15.38	8.66
	Native Hawaiian/Pacific Islander	23	16.83	6.76

Appendix H

2015 Performance Level Classifications

Note that these tables exclude those students who refused to test or were absent during some or most of the testing period; consequently, the rows across the five levels do not sum to 100%.

Table H.1
Percentage of Students at Each Performance Level 2015
Cluster 1, Kindergarten

Domain	Level 1	Level 2	Level 3	Level 4	Level 5
Reading	5.79	23.73	54.74	10.68	5.01
Writing	12.78	39.83	27.80	16.14	3.39
Listening	4.50	20.49	45.50	17.07	12.39
Speaking	6.15	19.56	37.04	22.77	14.43
Comprehension	5.79	23.76	54.71	11.01	4.68
Composite	12.09	39.92	35.09	10.92	1.92

Table H.2
Percentage of Students at Each Performance Level 2015
Cluster 2, Grades 1–2

Domain	Level 1	Level 2	Level 3	Level 4	Level 5
Reading	5.27	20.30	30.68	30.88	12.71
Writing	5.13	23.62	28.68	35.90	6.52
Listening	4.55	15.16	34.54	23.15	22.43
Speaking	5.34	10.41	28.39	31.10	24.60
Comprehension	5.25	20.19	31.10	30.97	12.33
Composite	6.33	25.43	33.23	30.19	4.66

Table H.3
Percentage of Students at Each Performance Level 2015
Cluster 3, Grades 3–5

Domain	Level 1	Level 2	Level 3	Level 4	Level 5
Reading	15.09	20.19	20.57	25.60	18.41
Writing	7.35	34.55	32.05	25.01	0.91
Listening	7.08	14.03	25.20	29.84	23.72
Speaking	3.81	2.30	14.67	30.11	48.97
Comprehension	13.80	20.50	22.54	26.19	16.84
Composite	13.30	27.64	31.75	26.48	0.70

Table H.4
Percentage of Students at Each Performance Level 2015
Cluster 4, Grades 6–8

Domain	Level 1	Level 2	Level 3	Level 4	Level 5
Reading	19.16	22.69	16.01	31.78	9.87
Writing	18.07	19.55	19.25	40.62	2.01
Listening	14.83	10.36	20.04	29.86	24.41
Speaking	7.66	6.88	11.15	25.93	47.89
Comprehension	18.37	22.54	17.73	31.53	9.33
Composite	18.61	22.50	25.93	31.83	0.64

Table H.5
Percentage of Students at Each Performance Level 2015
Cluster 5, Grades 9–12

Domain	Level 1	Level 2	Level 3	Level 4	Level 5
Reading	45.97	17.07	13.57	15.45	5.72
Writing	18.65	31.67	22.32	17.76	7.38
Listening	30.52	14.26	19.21	19.76	14.04
Speaking	20.70	9.26	14.21	11.48	42.13
Comprehension	44.81	17.67	14.17	15.49	5.63
Composite	39.69	21.94	18.48	15.24	2.43