Comprehensive
English
Language
Learning
Assessment

TECHNICAL SUMMARY REPORT

Developed by Educational Testing Service (ETS)
2005
Princeton, NJ

Reissued by AccountabilityWorks (AW)
2010
Bethesda, MD

Table of Contents

Section 1: Overview

The *Comprehensive English Language Learning Assessment* (CELLA) is a four-modality English language proficiency assessment designed to provide:

1. Evidence of program accountability in accordance with Title III of No Child Left Behind (NCLB), which calls for schools and districts to meet state accountability objectives for increasing the English-language proficiency of English Language Learners (ELL).
2. Data useful for charting student progress over time and, for newly arrived students, charting progress over the first year.
3. Information about the language proficiency levels of individual students that can be used in making decisions regarding placement into, or exit from, ESL or bilingual education programs.
4. Diagnostically useful information about individual students' strengths and weaknesses in English (with as much specificity as possible within the limitations of a large-scale standardized test).

Initial development of CELLA was funded by a grant from the U.S. Department of Education. The test was developed by ETS in collaboration with AccountabilityWorks and a consortium of five states: Florida, Maryland, Michigan, Pennsylvania, and Tennessee. The test items included in this assessment are based on the CELLA assessment benchmarks that, in turn, are aligned to the English language development standards of the five consortium states.

This document summarizes the field testing, calibration, and vertical scaling of items used to develop two final test forms for each of four test levels as well as a locator test. Each test level is associated with specific grades and provides assessment in four modalities: Listening, Speaking, Reading, and Writing. Listening and Speaking items were calibrated together on a single vertical scale. In addition, vertical scales for Reading and for Writing were created.

The final forms associated with the four levels of the operational test in the modalities Reading and Writing are designed to be administered to the following grade levels:

- Grades K–2: Level A
- Grades 3–5: Levels A, B
- Grades 6–8: Levels A, B, C
- Grades 9–12: Levels A, B, C, D

The CELLA system implements functional level testing in Reading and Writing based on the results of a brief locator test (to be used primarily with students taking CELLA for the first time—i.e., for whom there is no prior CELLA assessment data). This feature means that an ELL student could be administered any of the four Levels A, B, C, or D depending on his/her level of functioning in English. For example, a high school student in grade 10 could legitimately be administered the Level C test if use of the Level C test would provide a better measure of his/her English skills in Reading and Writing. (Development and appropriate use of the vertical scales and the locator test are described in *Section 7: Vertical Scaling and Section 9: Final Test Forms and the Locator Test*.) The four CELLA levels in Reading and Writing address skills typically required in regular instructional environments (i.e., non-Limited English Proficient or non-LEP) at the following grade levels: Level A (K–2); Level B (3–5); Level C (6–8); Level D (9–12).

The final forms associated with the Speaking and Listening modalities are designed to be administered at the following grade levels:

- Grades K–2: Level A
- Grades 3–5: Level B
- Grades 6–8: Level C
- Grades 9–12: Level D

There is not the same need to implement functional level testing in these modalities because there is greater content range in these CELLA subtests and because LEP students tend to develop oral language skills at a substantially faster pace than written language skills.

The field test was conducted in the Fall of 2004. Fall performance is more similar to performance in the Spring of the previous grade than to Spring of the same grade; for example, students in the Fall of grade 6 are more similar to students in the Spring of grade 5 than to students in the Spring of grade 6. Accordingly, for the field test, the test levels and their associated grades were as follows:

- Field Test Level A (administered at grades K–3)
- Field Test Level B (administered at grades 4–6)
- Field Test Level C (administered at grades 7–9)
- Field Test Level D (administered at grades 10–12)

The operational CELLA offers users the option of administering the Reading and Writing levels of the CELLA test at grade levels based on the regular instruction (non-LEP) content equivalents: Level A (K–2); Level B (3–5); Level C (6–8); Level D (9–12).

Section 2: Field Test

The field test took place between October 25 and November 8, 2004. This testing window was extended slightly in some instances, because a few schools requested later testing times.

Field Test Participants

Students who participated in the field test came from the five states in the consortium: Florida, Maryland, Michigan, Pennsylvania, and Tennessee. Students in kindergarten through grade 12 were involved.

Each state in the consortium selected field test participants from the population of students in each state who might take CELLA in an operational administration. Each state was asked to select students who reflected the most prevalent foreign languages spoken by ESL students in the state. Also, each state was asked to select students from the full range of proficiency levels.

ETS gave each state in the consortium a list of school districts with a designated number of students per grade level to be recruited. This list was gathered from a national database that identified districts with programs for English Language Learners, described the size of the district, and designated whether the district was classified as rural, urban, or suburban. State coordinators reviewed the list to evaluate whether this list was likely to yield a sample that included the most prevalent languages spoken by ESL

students and would cover the range of proficiency levels of ESL students in the state. Revisions in the list were made where necessary.

A state coordinator worked with the testing coordinator of each district on the final list to select the schools and classrooms/students to participate in the field test. Within each designated school, the district testing coordinator and school principal determined whether to sample intact classrooms or sample individual students. Where possible, existing data describing students' levels of English language proficiency were considered as part of the process of selecting the classrooms/students to participate. Some students included in the field test were those who had been recently reclassified as proficient in English. Table 1 shows the number of students from each state that participated in the field test.

It should be noted that the students included in the CELLA field test were not drawn from the same populations in all grade levels; some schools participating in the study contributed students at only two grade levels, whereas others contributed students at five grade levels. Therefore, at adjacent grade levels there were likely to be some differences in the abilities of the samples tested.

The Tests Administered

Table 2 describes the composition of the field test forms that were administered. As noted previously, Level A field test forms were administered to students in kindergarten through grade 3. Grade 3 students were administered additional Reading and Writing items targeted at their level. (The test forms containing the Level A items plus these additional items were called Level A Extension forms.) Level B field test forms were administered in grades 4, 5, and 6, and Level C forms were administered in grades 7, 8, and 9. Finally, Level D forms were given in grades 10, 11, and 12.

The Listening section consisted of multiple-choice (MC) items and constructed-response (CR) items at Level A; the other Listening levels contained MC items only. The Speaking test consisted of CR items that were individually administered and scored locally by classroom teachers. At Levels A and B, the Reading sections contained CR items and MC items; Levels C and D Reading contained MC items only. Finally, the Level A and A Extension Writing tests contained only CR items, while the Writing tests at Levels B, C, and D contained both MC and CR items. Appendix B provides more details about the skills assessed by each modality. (For consistency with terminology in other CELLA assessment design documents, the term "benchmarks" is used to describe the major skill categories.) Please note that the test blueprint in this appendix refer to the final operational test forms, not the field test forms.

There were three field test forms at each level, from which appropriate items were selected to construct two operational forms. Table 2 shows that the test forms associated with each test level contained unique items, as well as common items that were shared across the other test forms and used to link the forms horizontally. In the three Level A Listening sections, for example, there were 8 MC items and 3 CR items that were unique, which means that these items were not shared across the three forms. In addition there were 7 MC items and 2 CR items that were common to the three forms, which means that all three forms contained these items. These common items were used to link the Level A test forms horizontally. More information about this linking is provided in *Section 6: Item Calibration*.

In addition, Table 2 shows that all three Level A forms of the Listening test contained MC vertical scaling items that were "Level Up." These items were the common items in the Level B Listening tests. The A-to-B vertical linking items for Level B were drawn from the common items in the Level A Listening tests. The function of these vertical linking items was to enable a link to be made between the Level A

and Level B Listening tests. More details about the vertical scaling procedures are provided in *Section 7: Vertical Scaling*.

The Level B test forms in Table 2 that are called B4, B5, and B6 merit a brief discussion. The on-level items contained in these forms were the same items as those appearing in B1, B2, and B3, respectively; only the vertical linking items differed. For example, Table 2 shows that forms B1, B2, and B3 contained a set of vertical linking items from Level A, whereas forms B4, B5, and B6 contained a set of vertical linking items from Level C. Two sets of forms were prepared so that students would respond to only one set of linking items. The six Level C test forms were configured similarly.

For Writing forms B1, B2, and B3 there were no Level A linking items. The link between these levels was based on Level B items in the A Extension forms.

Test Administration Procedures

At each grade level, test forms were spiraled over classrooms or test groups. Students who were administered Level A and A Extension test forms recorded their answers in their test booklets. Students administered test levels B, C, and D recorded their answers on a separate answer sheet. Level A students taking Level B linking items marked their responses in their test booklets; Level B students taking the Level A linking items marked their responses on the separate answer sheet. All test items taken by kindergarten and grade 1 students were individually administered. At all grade levels, the Speaking section was individually administered and teacher scored, with scores recorded on the student answer sheet.

In the field test, the Speaking section also included a pronunciation score. This score did not correspond to any single item; rather it reflected the administrator's judgment of the student's pronunciation based on all items in the Speaking section. The pronunciation score was calibrated, but it was later dropped from the item pool during the data review process.

Testing times varied as a function of test level and the number of individually administered items, but approximate times were: Level A test forms—1 hour and 35 minutes; Level A Extension forms—2 hours and 20 minutes; Level B, C, and D test forms—3 hours and 45 minutes. Test administrators were encouraged to give students additional time if needed.

Section 3: Item Analyses

Item Analyses (IA) were conducted to establish the reasonableness of keys and the reliability of each field test. Summary statistics describing the difficulty, discrimination, and internal consistency of the on-level items that were field tested are given in Table 3. To assess the difficulty of each multiple-choice item, the proportion of students correctly answering the item (called the p-value) was used as the index of item difficulty; for constructed-response items an analogous index consisting of the mean item score divided by the maximum possible item score was used. To assess discrimination, the correlation between students' item scores and their total test scores was used for both item types. To evaluate internal consistency, Cronbach's α was used for the Listening and Reading test sections while a stratified coefficient alpha was used for the Writing and Speaking test sections.

Table 3 provides summary statistics that describe the items in each field test form. In this table, results are combined across grade levels. The results in Table 3 indicate that the three test forms associated with each test level within each modality were very similar in terms of difficulty and discrimination.

The p-values for the Listening items generally were in the low 0.70s on average and their item-test correlations near 0.50 on average. The p-values for the Speaking items also tended to be in the low 0.70s on average, but they had notably higher item-test correlations, probably because there was no guessing on these individually administered items. In general, the Reading items were notably harder than the Listening and Speaking items but similar in item-test correlation, except for the Level A Extension items, which had p-values that were higher and item-test correlations that were slightly lower on average, probably because of the additional MC items they contained. With respect to Writing, the Level A items were very difficult and very discriminating. In the remaining test levels, the items were notably easier and somewhat less highly correlated with the total test score.

Section 4: Raw Score Summary Statistics

Table 4 provides raw score summary statistics by modality, grade, and form for the on-level field test items. Tables 5, 6, and 7 contain cumulative raw score distributions, pooled across forms, for Listening/Speaking, Reading, and Writing, respectively.

Average raw scores generally increase across grades of students administered the same test level, but this is not universally the case. These results are not unexpected for the following reasons. When traditional achievement tests for native-English speakers are administered to samples of students that are controlled to be consistent samples of the same population across grades, scores typically increase as grade level increases. Growth decreases as grade increases, and at high school, growth can be minimal or there can be decreasing scores (probably due to declining student motivation).

With respect to student scores on the CELLA field test forms, expectations of score change over grades are not as clear-cut. First, the field test samples were in the range of about 350 to 425 students per grade and form; these samples were quite adequate for the scaling process when pooled across grades, but they are relatively small for estimating population characteristics by grade. Further, the students included in the CELLA field test were not drawn from the same populations in all grade levels; some schools participating in the study contributed students at only two grade levels, whereas others contributed students at five grade levels. Therefore, at adjacent grade levels there were likely to be differences in the abilities of the samples tested. In addition, there can be immigration trends (e.g., if there are influxes of students to the United States in order to attend high school) that produce expectations of lower scores at some grades. Finally, for tests such as Listening and Speaking, which assess skills learned outside as well as in school, it is not clearly known what bearing grade level has on the acquisition of these skills.

Section 5: Evaluation of Differential Item Functioning

Following the classical item and raw score analyses, a Differential Item Functioning (DIF) study was carried out. One of the goals of test development is to assemble a set of items that provides an estimate of a student's ability that is as fair and accurate as possible for all groups within the population. DIF statistics are used to identify items that function differently for particular subgroups of students (e.g., females versus males); this differential item functioning occurs when students in different subgroups who have the same underlying level of ability have different probabilities of answering the item correctly.

An item that is found to be easier or harder for different subgroups of students having the same ability is flagged as an item that has DIF. Items that are flagged are subsequently reviewed by content experts to see if the experts can identify the source and meaning of the DIF. One possible source is bias, but it is also possible that DIF can occur because there are real differences between the subgroup examinees in the knowledge or skill assessed by a flagged item. Also, it is possible that the flagging is due to a statistical Type I error.

ETS used two statistical DIF detection methods, the Mantel-Haenszel (Mantel & Haenszel, 1959) and the Standardization (Dorans & Holland, 1993) approaches. As part of the Mantel-Haenszel procedure, the statistic described by Holland & Thayer (1988), known as MH D-DIF, is used. This statistic captures the differences between two groups, called the focal and reference groups, after conditioning on total test score.

The standardized mean difference (SMD) is used as an index of severity in conjunction with the MH statistic. The SMD compares the item means of the two studied groups after adjusting for differences in the distribution of members across the values of the matching variable (total test score).

The evaluation of DIF is a two-step procedure that combines the results of the statistical test and the effect size. First, the MH statistic and the SMD are computed for each item. Second, each item is classified into different categories based on its MH statistic and corresponding SMD value. Items that are not statistically significant based on the MH statistic ($p>0.01$) are considered to have similar performance between the two studied groups; these items are considered to be functioning appropriately. For items where the statistical test indicates significant differences ($p<0.01$), the SMD is used to determine the direction and severity of the DIF.

Items are classified into one of three categories and assigned values of A, B, or C. Category A items contain negligible DIF, Category B items exhibit slight or moderate DIF, and Category C items have moderate to large values of DIF. For constructed-response items the MH D-DIF is not calculated, but analogous flagging rules based on the chi-square statistic were applied, resulting in classification into A, B, or C DIF categories.

The DIF analyses carried out for CELLA consisted of comparisons between the item performances of male students versus female students. In total, five items were flagged for DIF. These items were reviewed by content experts to determine if the items contained inappropriate content. Three of the items were dropped from the pool, and two of the items were retained in the item pool and used in the final test forms that were subsequently developed.

Section 6: Item Calibration

The items in the field test forms were calibrated using a combination of item response theory models. The three-parameter logistic model was used for multiple-choice items and the generalized partial credit model was used for the constructed-response items. All items were calibrated using the PARSCALE component of ETS' proprietary software, GENASYS.

There were three separate sets of calibrations: Listening/Speaking, Reading, and Writing. The calibrations were carried out by test level, as is shown in Table 8. All forms associated with a given test level were calibrated simultaneously. The common items shared by the forms within a test level served to link the forms so that the resulting parameters for the items in all of the forms calibrated together were expressed on the same scale. All items were successfully calibrated, and none were found to have poor fit.

Section 7: Vertical Scaling

The items calibrated by test level were placed on a common vertical scale using the test characteristic curve (TCC) method described by Stocking and Lord (1983). This procedure examines the true scores obtained for a set of items that has parameters on both a reference scale and an anchor scale, and finds the linear transformation of the anchor scale item parameters that minimizes the sum of squared differences between the TCCs. This linear transformation is used to transform the item and ability parameters for the test containing the anchor item set so that these parameters are expressed on the same score scale used by the reference item set.

The series of analyses conducted for the vertical scaling is shown in Table 9. The Level B calibrations were chosen to define the base scale of measurement for Listening/Speaking, Reading, and Writing. The forms associated with the other test levels were placed on this scale in sequence. For Listening/Speaking, the Level A calibrations were linked to the Level B scale using the TCC method based on the Level B items embedded in test forms A1, A2, and A3, and the Level A items embedded in test forms B1, B2, and B3. Then, the Level C calibrations were linked to the Level B scale via the Level C items embedded in test forms B4, B5, and B6 and the Level B items embedded in test forms C1, C2, and C3. Finally, the Level D calibrations were linked to the common scale now comprised of Level A, B, and C items via the Level D items included in forms C4, C5, and C6 and the Level C items embedded in test forms D1, D2, and D3. In the case of Reading and Writing, the Level A vertical scaling items used to link Levels A and B were in the Extension forms.

The vertically linked test forms produced parameters expressed in the theta metric. The results on the theta metric were transformed to a 3-digit scale score metric. This transformation enabled scores to be reported in a metric more familiar and comprehensible to typical test users. The particular values of the transformation were selected to produce a scale on which the score distribution for students in the field test who took Level B (excluding those students with scores at or below the chance level or with perfect scores) would have a mean of approximately 700 and a standard deviation of approximately 40. Finally, the lowest obtainable scale scores (LOSS) and highest obtainable scale scores (HOSS) were set for scores at or below the chance level and perfect scores, respectively. The Listening/Speaking scale scores range from a low of 495 for the LOSS for Level A to a high of 835 for the HOSS of Level D. The Reading scale ranges from 345 to 820, and the Writing scale ranges from 515 to 850.

The following paragraphs present information about the vertical scales in terms of test characteristic curves (TCC) and conditional standard errors of measurement (SEMs) for the field test forms.

Figures 1, 2, and 3 display the TCCs for the three scales for the field test forms for Listening/Speaking, Reading, and Writing, respectively. The TCCs show expected proportion-correct scores at each scale score. In general TCCs for alternate forms at the same level tend to be similar. TCCs for higher levels are generally expected to move up the scale—i.e., it is generally expected that students with a given scale score will have a lower expected proportion correct on a higher test level. However, it is not uncommon for disordinalities between levels to occur, especially among higher test levels in the junior and senior high school grades. In general the CELLA field test levels do display ordinality. For Writing, Levels C and D are very similar in difficulty and this similarity carries through to the final test forms.

It is interesting to note that Reading Level A covers a wide range of scale scores, much wider than that seen with Listening/Speaking or Writing. This result reflects the fact that the Level A Reading test covers a very broad set of skills, ranging from identifying letters of the alphabet through reading a passage of several paragraphs and answering questions about it. Thus, students with very low Reading proficiency

can be expected to grow many scale score points during a school year, and as their proficiency increases beyond the very basic skills, expected growth will be much smaller in magnitude. In this case, caution is needed when interpreting growth on Reading Level A as compared to the other Reading levels.

The conditional SEMs of the field test forms appear in Figures 4, 5, and 6 for Listening/Speaking, Reading, and Writing, respectively.

Table 10 provides summary statistics that describe students' field test scale scores by grade. Results are pooled over forms. Cumulative scale score distributions appear in Tables, 11, 12, and 13. Plots of the cumulative distributions of these scale scores are given in Figures 7, 8, and 9.

As discussed in *Section 4: Raw Score Summary Statistics*, traditional expectations of increasing scores as grade increases cannot necessarily be expected to apply to an English-acquisition assessment or to the data in this field test. It is important to note that when students in successive grades have taken the same test level, disordinality between grades must be attributed to student or skill differences, because the vertical scaling does not affect those results. Vertical scaling can only affect the alignment between test levels, not the ordering of scores within test level.

For the most part, scores on the field tests tended to increase as grade increased, but this was not a universal finding. For example, high school Writing scores did not increase as grade increased, even though those students all took the same test level. Mean scores were particularly susceptible to effects of the proportions of students scoring at the floor (LOSS) of the tests. The Level A Reading and Writing field test forms were particularly difficult for students in the Fall of kindergarten, where 55% to 51% of the students performed at the floor of the test.

As with any vertical scale, care should be used in comparing scale scores that are provided by test levels that differ greatly in difficulty and include substantially different types of test questions. For example, as noted above, Level A Reading scale scores are likely to provide scale score growth and variances that differ from those provided by other Reading test levels.

Section 8: Scale Anchoring

In order to increase the interpretability of the CELLA scale scores, scale anchoring was conducted. This process identified selected scale score anchor values throughout the range of performance. Students who performed near these anchor values on the field test were identified. Based on the performance of these students, items were identified on which these students usually were successful. Content experts reviewed these items and developed behavioral descriptions that highlighted the major behaviors typically evidenced by students at each anchor value. These descriptions can help students, parent, and teachers understand the meaning of student scale scores.[1]

The technical details of the scale anchoring process are described in Appendix C.[2] The Listening/Speaking, Reading, and Writing scales were all designed to be centered at Level B with an average score of approximately 700 and a standard deviation of approximately 40 (when students at the

[1] The scale anchoring process differs from a standard setting process and does not replace it. In standard setting, expert committees are given performance level descriptions (e.g., for Proficient or Advanced performance) and information about test items, and tasked with identifying test performance that is required, according to their judgment, to reach each performance level.
[2] In the Appendix, the term "benchmark" is used. In order to reduce confusion with other uses of that term, the scale anchoring is described here in terms of "anchor points" rather than "benchmarks."

LOSS and HOSS were not included). Based on the general procedure described in Appendix C, and given the characteristics of the three CELLA scales, scale score anchor points were chosen to be at:

Anchor Point 1: 620=700–2x40
Anchor Point 2: 660=700–40
Anchor Point 3: 700
Anchor Point 4: 740=700+40

For Reading and Writing, data were available to define one additional anchor point:

Anchor Point 5: 780=700+2x40

(For Listening/Speaking, there were insufficient numbers of items located near 780 and insufficient numbers of students performing at 780 to define this additional anchor point.)

At each anchor point, students were identified who scored within +/-10 scale score points of the anchor point. Item performance for those students was evaluated. At each anchor point, multiple-choice items were identified for which (a) these students had a proportion correct ≥ 0.74, and (b) students at the next lower anchor point had a proportion correct of at least .25 lower. Score levels for CR items were identified for which (a) the anchor point students had an expected probability of obtaining that score level or higher of at least 0.65, and (b) students at the next lower anchor point had at least a .25 lower expected probability of obtaining that score level or higher. For a CR item, one score level might qualify for a lower anchor point, and a higher score level from that same item might qualify for a higher anchor point.

Exemplar items and item score levels could come from any field test level contributing to the behavioral description for a particular anchor point. For lower anchor points, most items came from Levels A and B; for the highest anchor points, most items came from Levels C and D. The content experts who developed the behavioral descriptions typically worked with about 20-30 exemplar items per anchor point. If more items than this were available statistically, items were randomly selected to provide the necessary 20–30 items.

The content experts carefully reviewed the exemplar items at each scale anchor point and synthesized the common knowledge or skills that were typically evidenced by students at each anchor point. In addition, the content experts compared the exemplar items in adjacent anchor points to differentiate the level of skills they required. This process was used repeatedly for each vertical scale to capture the progression of skills across the scale.

It should be kept in mind when using the behavioral descriptions that they are conceptual descriptions of typical performance. The actual behaviors evidenced by any given student may vary from these descriptions for three reasons. First, every test score contains measurement error, and students are likely to obtain somewhat different scores on different testing occasions. Second, students who receive the same scale score on the same test form can do so with different patterns of skills. For example, a student who obtains a raw score of 21 (out of 34 possible) on final Reading form A1 Extension will obtain a scale score of 645. Students can obtain a raw score of 21 by getting different types of items correct—not all students who obtain a score of 21 are exactly the same. Third, students who take different test levels interact with different item types. In particular, it should be noted that a small number of Level A students' scale scores may reach Reading and Writing anchor points 4 or 5. These students share characteristics with Level C and D students, but they have not read passages nor answered items that have the same complexity as those appearing in Levels C and D. Caution should be used in interpreting

Level A students' scale scores that reach these high anchor points. Thus, for the reasons presented, the behavioral descriptions for all the scales are generalities that describe typical performance at each anchor point, not necessarily exactly what a given student can do.

The behavioral descriptions for the anchor points are described on the following pages. The percents of students in the field test who reached each anchor point are described in Tables 14, 15, and 16.

LISTENING/SPEAKING

Anchor Point 1 (620)
Students scoring at this point on the scale typically. . .

- demonstrate very limited vocabulary and grammar resources; they can use some basic words but have difficulty producing and understanding basic questions and taking part in even simple exchanges.

Anchor Point 2 (660)
Students scoring at this point on the scale typically. . .

- demonstrate limited vocabulary and grammatical resources; they have command over a basic vocabulary; they can produce and understand basic exchanges, though their limited proficiency often impedes communication.

Anchor Point 3 (700)
Students scoring at this point on the scale typically. . .

- demonstrate adequate vocabulary and grammatical resources; they can communicate effectively (though with some errors) in everyday situations and can generally understand and participate in many classroom activities (e.g., following directions and announcements). They are developing the ability to comprehend grade-level instruction in the content areas; they can do it sometimes, but not consistently.

Anchor Point 4 (740)
Students scoring at this point on the scale typically. . .

- demonstrate a solid command of grammar and vocabulary, including common idioms. They can communicate effectively in everyday and academic situations, can understand and participate in most classroom activities, and are able to use English to learn new information in the content areas. They may speak with an accent and make occasional errors, but they are able to communicate effectively on a range of topics.

READING

Anchor Point 1 (620)
Students scoring at this point on the scale typically. . .

- demonstrate that they are just beginning to read. They understand basic concepts of print (e.g., recognize the directionality of English print), decode short words, and can read most common sight words. They can also read very simple single sentences and respond to questions about their meaning.

Anchor Point 2 (660)
Students scoring at this point on the scale typically. . .

- demonstrate that they are still in the process of "learning to read" but are on the verge of transitioning to being able to "read to learn." They can independently read short passages written in very simple language on a range of topics. They can answer simple, explicit main idea questions and simple, literal detail questions related to these passages.

Anchor Point 3 (700)
Students scoring at this point on the scale typically. . .

- demonstrate that they are developing as independent readers. They can read short passages written in very simple language with complete fluency and can answer any type of question about such simple passages, including inference questions and questions about characters' feelings. They can read short passages of moderate complexity, with partial comprehension; they can answer literal main idea and detail questions related to such passages, but may struggle with questions requiring inference or interpretation.

Anchor Point 4 (740)
Students scoring at this point on the scale typically. . .

- demonstrate that they are developing as independent readers of challenging texts, with adequate vocabulary resources. They can read short passages of moderate complexity with thorough comprehension. When reading short passages that are more complex, they have only partial comprehension. They can, however, answer a range of questions related to such passages, including questions about sequence of events and text organization and questions requiring simple inferences.

Anchor Point 5 (780)
Students scoring at this point on the scale typically. . .

- demonstrate that they are developing as independent readers of the most challenging of texts. They can read the most linguistically complex short passages with good comprehension. Drawing on well-developed vocabulary resources and syntactic knowledge, they are able to distinguish subtle differences in meaning. They can answer a range of questions that require synthesizing information, making inferences, identifying the important details, and identifying a main idea when it is not explicitly stated.

WRITING

Anchor Point 1 (620)
Students scoring at this point on the scale typically. . .

- demonstrate a developing knowledge of sound/spelling relationships; they are generally able to write dictated letters and words, though with occasional errors.

Anchor Point 2 (660)
Students scoring at this point on the scale typically. . .

- demonstrate a solid command of sound/spelling relationships; they can accurately write dictated letters and words. They are beginning to develop the ability to write original descriptive and interrogative sentences and use capital letters and punctuation.

Anchor Point 3 (700)
Students scoring at this point on the scale typically. . .

- demonstrate basic vocabulary resources and a partial control of grammar and the conventions of written English. They can write original descriptive and interrogative sentences as well as narrative and descriptive paragraphs, though their writing contains significant and/or numerous errors that may interfere with communication.

Anchor Point 4 (740)
Students scoring at this point on the scale typically. . .

- demonstrate adequate vocabulary resources and an adequate, if imperfect, command of grammar and the conventions of written English. They can write original descriptive and interrogative sentences as well as narrative, descriptive, and personal opinion paragraphs; their writing is accurate enough to communicate effectively but may contain errors or be marked by simple structures used to avoid errors. They are still developing the ability to write paragraphs in more challenging genres such as comparison and contrast.

Anchor Point 5 (780)
Students scoring at this point on the scale typically. . .

- demonstrate well-developed vocabulary resources and excellent control of grammar and the conventions of written English. They can write paragraphs of a range of genres (e.g., descriptive, persuasive, compare and contrast) that are well developed and marked by advanced grammatical structures.

Section 9: Final Test Forms and the Locator Test

Final Test Forms

Final forms were created by attempting to keep the highest quality field test forms intact. Items were swapped in and out of the intact form only as necessary to balance the difficulty and content representation of item types. Items for the second final form at each level were selected so that the difficulty and discrimination of the items matched those of the first final form. An effort also was made to keep the operational item order very close to what it was in the field test form. Items were sequenced within form from easy to hard, when possible. Both final forms were assembled according to the test blueprint.

Table 17 contains the means and standard deviations of the item difficulties and item discriminations for the final test forms.

Scoring tables for these final test forms are available in a separate document. Please contact Accountability Works/ETS for information on final scoring tables.

Locator Test

A Locator Test was developed for use in conjunction with functional level testing for the Reading and Writing sections. Functional level testing allows students in grades 3-12 to take a lower level of the Reading and Writing sections if their literacy skills are at a more basic level.

Locator Test Construction and Scaling

The Locator Test contains multiple-choice Reading items. These items were selected from among the items that were field tested but not chosen for one of the operational test forms. Representative items of good quality were chosen. Three items were selected from Level A Extension forms, 5 items from Level B forms, 5 from Level C forms, and 5 from Level D forms, for a total of 18 items.

The selected items had item parameters calibrated on the vertical scale for Reading. The Raw Score-to-Scale Score scoring table based on those parameters appears in Table 18. This table also includes the conditional SEMs for each scale score value obtainable on the Locator Test.

Selection of Cut Points

The Locator Test's primary function is to determine whether the Reading and Writing sections of a given level of CELLA will be too difficult for a student, so that it would be advisable to administer a lower level of these sections to that student. Two major criteria were used to select cut points between levels: (a) students needed to have a low expected proportion-correct score (approximately in the .35 to .40 range) on the upper of the two levels so that the recommendation of the lower (easier) level would be appropriate, and (b) the vast majority of students (approximately 75 to 80%) of students who took the upper level during the field test scored above the cut point. Using these criteria, the scale score values available for the Locator Test were considered for the cut points and the following determinations were made.

Raw Score on Locator Test	Scale Score on Locator Test	Recommended Test Level
0–5	345	Level A
6–8	652–692	Level B
9–12	702–723	Level C
13–18	729–820	Level D

The expected proportion-correct score on Level B for students with a scale score of 652 was 0.33; the expected proportion-correct score on Level C for students with a scale score of 702 was 0.40; and the expected proportion-correct score on Level D for students with a scale score of 729 was 0.42. Table 19 shows the proportion of students in the field test taking each test level scoring below, versus at or above, each relevant cut point. Eighty-three percent of the students taking Level B scored at or above 652, 74% of the students taking Level C scored at or above 702, and 77% of the students taking Level D scored at or above 729.

The CELLA Locator Test is a tool to be used as part of the decision-making process in determining whether or not functional level testing of the Reading and Writing sections should occur. This information should be used along with teacher judgment, classroom performance, and all other available relevant information in deciding what level of the CELLA Reading and Writing sections a student should take. More information about the use of the Locator Test is contained in Appendix D.

Section 10: References

Dorans, N.J., & Holland, P.W. (1993). DIF Detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Ed.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719–748.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7(2)*, 201–210.

Appendix A: Tables and Figures

Table 1

Number of Students in the Field Test
by State and Grade Range

State	Grade Range			
	K, 1, 2, 3	4, 5, 6	7, 8, 9	10, 11, 12
Florida	1348	1003	1035	945
Maryland	1273	835	942	750
Michigan	859	691	711	640
Pennsylvania	1233	749	709	585
Tennessee	711	444	369	349
Total	5424	3722	3766	3269

Table 2

Configuration of Field Test Forms

Content Area	Test Level	Forms	Number of Items by Item Type									
			On-Level Items Being Field Tested				Vertical Linking Items				Total	
			Unique		Common		Level Up*		Level Down*			
			MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
Listening	A	A1, A2, A3	8	3	7	2	15				30	5
	B	B1, B2, B3	13		15				7		35	
	B	B4, B5, B6[1]	13		15		15				43	
	C	C1, C2, C3	13		15				15		43	
	C	C4, C5, C6[1]	13		15		15				43	
	D	D1, D2, D3	13		15				15		43	
Speaking	A	A1, A2, A3		6		5		6				17
	B	B1, B2, B3		8		6				5		19
	B	B4, B5, B6[1]		8		6		6				20
	C	C1, C2, C3		8		6				6		20
	C	C4, C5, C6[1]		8		6		6				20
	D	D1, D2, D3		8		6				6		20
Reading	A	A1, A2, A3	9	4	6	2					15	6
	A	A1+Ext, A2+Ext, A3+Ext	16	4	9	2	12				37	6
	B	B1, B2, B3	18	1	12				9		39	1
	B	B4, B5, B6[1]	18	1	12		14				44	1
	C	C1, C2, C3	20		14				12		46	
	C	C4, C5, C6[1]	20		14		12				46	
	D	D1, D2, D3	21		12				14		47	
Writing	A	A1, A2, A3		4		3						7
	A	A1+Ext, A2+Ext, A3+Ext		7		9	8	3			8	19
	B	B1, B2, B3	11	3	8	3					19	6
	B	B4, B5, B6[1]	11	3	8	3	8	3			27	9
	C	C1, C2, C3	11	3	8	3			8	3	27	9
	C	C4, C5, C6[1]	11	3	8	3	8	3			27	9
	D	D1, D2, D3	11	3	8	3			8	3	27	9

[1] Forms 4, 5, and 6 contain the same on-level items as do Forms 1, 2, 3, respectively, but they contain different sets of vertical linking items.

* 'Level Up' and 'Level Down' refers to items from one level above and one level below respectively.

Table 3

Item Statistics by Modality, Test Level, and Field Test Form

			ALL items						MC items					CR items				
				Difficulty		Discrimination				Difficulty		Discrimination			Difficulty		Discrimination	
Modality	Test Level/ Form	No. of Cases	No. of Items	Mean	SD	Mean	SD	Internal Consistency	No. of Items	Mean	SD	Mean	SD	No. of Items	Mean	SD	Mean	SD
Listening	A1	1842	20	0.70	0.17	0.53	0.11	0.81	15	0.64	0.15	0.54	0.11	5	0.86	0.11	0.51	0.12
	A2	1917	20	0.72	0.17	0.52	0.11	0.79	15	0.65	0.15	0.56	0.10	5	0.92	0.04	0.40	0.04
	A3	1427	20	0.74	0.17	0.50	0.12	0.78	15	0.68	0.16	0.53	0.11	5	0.90	0.06	0.40	0.10
(All MC items	B1	1206	28	0.70	0.17	0.49	0.08	0.87										
for Levels	B2	1116	28	0.71	0.12	0.49	0.07	0.87										
B, C, & D)	B3	1231	28	0.69	0.14	0.47	0.08	0.84										
	C1	1034	28	0.70	0.12	0.54	0.07	0.90										
	C2	1161	28	0.71	0.11	0.52	0.08	0.88										
	C3	1243	28	0.74	0.12	0.52	0.08	0.89										
	D1	989	28	0.77	0.15	0.47	0.10	0.85										
	D2	853	28	0.73	0.16	0.46	0.08	0.84										
	D3	1072	28	0.76	0.13	0.45	0.08	0.82										
Speaking	A1	1842	11	0.71	0.14	0.69	0.09	0.87										
(All CR items)	A2	1917	11	0.69	0.13	0.66	0.13	0.85										
	A3	1427	11	0.73	0.14	0.70	0.11	0.88										
	B1	1206	13	0.76	0.12	0.67	0.16	0.93										
	B2	1116	13	0.78	0.12	0.72	0.15	0.93										
	B3	1231	13	0.74	0.10	0.70	0.12	0.92										
	C1	1034	13	0.68	0.11	0.70	0.17	0.94										
	C2	1161	13	0.68	0.08	0.73	0.15	0.95										
	C3	1243	13	0.65	0.13	0.68	0.17	0.94										
	D1	989	13	0.76	0.10	0.64	0.18	0.92										
	D2	853	13	0.75	0.08	0.67	0.17	0.92										
	D3	1072	13	0.75	0.11	0.62	0.22	0.91										

Table 3 (cont.)

Item Statistics by Modality, Test Level, and Field Test Form

			All Items						MC items					CR items				
				Difficulty		Discrimination				Difficulty		Discrimination			Difficulty		Discrimination	
Modality	Test Level / Form	No. of Cases	No. of Items	Mean	SD	Mean	SD	Internal Consistency	No. of Items	Mean	SD	Mean	SD	No. of Items	Mean	SD	Mean	SD
Reading	A1	1263	21	0.56	0.19	0.57	0.13	0.81	15	0.49	0.18	0.55	0.14	6	0 .74	0 .05	0.62	0.09
	A2	1242	21	0.60	0.21	0.60	0.13	0.85	15	0.52	0.18	0.60	0.14	6	0 .80	0 .10	0.59	0.14
(All MC items	A3	900	21	0.59	0.19	0.56	0.15	0.82	15	0.53	0.19	0.54	0.16	6	0 .74	0 .08	0.61	0.11
for Levels	A1+Ext	517	31	0.76	0.19	0.43	0.11	0.82	25	0.73	0.19	0.43	0.10	6	0 .90	0 .15	0 .42	0.16
A, C, & D)	A2+Ext	561	31	0.75	0.21	0.40	0.11	0.76	25	0.72	0.21	0.43	0.07	6	0 .89	0 .12	0.28	0.18
	A3+Ext	429	31	0.74	0.19	0.42	0.11	0.77	25	0.71	0.19	0.40	0.10	6	0.87	0.14	0.48	0.15
	B1	1251	31	0.58	0.16	0.50	0.10	0.87	30	0.58	0.16	0.50	0.10	1*	0.53	-	0.62	-
	B2	1180	31	0.58	0.14	0.51	0.08	0.89	30	0.58	0.14	0.51	0.08	1*	0.51	-	0.70	-
	B3	1277	31	0.57	0.15	0.49	0.10	0.87	30	0.56	0.15	0.49	0.10	1*	0.51	-	0 .59	-
	C1	1096	34	0.54	0.13	0.48	0.09	0.88										
	C2	1190	29	0.53	0.13	0.47	0.08	0.86										
	C3	1249	34	0.57	0.15	0.47	0.11	0.88										
	D1	997	32	0.63	0.14	0.47	0.09	0.87										
	D2	967	33	0.61	0.14	0.49	0.08	0.89										
	D3	1117	33	0.60	0.13	0.47	0.09	0.88										
Writing	A1	1373	7	0 .38	0 .13	0 .85	0 .04	0 .94										
	A2	1366	7	0 .37	0 .15	0 .85	0 .06	0 .94										
(All MC items	A3	1092	7	0 .37	0 .15	0 .85	0 .06	0 .95										
for Level A)	A1+Ext	545	16	0.53	0.12	0.65	0.15	0.92	5	0.67	0.15	0.47	0.08	11	0.73	0.11	0.73	0.09
	A2+Ext	580	16	0.70	0.13	0.66	0.15	0.93	5	0.66	0.15	0.47	0.06	11	0.72	0.12	0.75	0.08
	A3+Ext	455	16	0.69	0.14	0.68	0.14	0.93	5	0.65	0.17	0.51	0.06	11	0.72	0.13	0.75	0.08
	B1	1255	25	0.58	0.11	0.55	0.18	0.94	19	0.57	0.11	0.46	0.10	6	0 .63	0.11	0.82	0.01
	B2	1182	25	0.59	0.12	0.54	0.18	0.94	19	0.58	0.13	0.46	0.10	6	0 .62	0.11	0.82	0.02
	B3	1285	25	0.59	0.12	0.54	0.18	0.94	19	0.57	0.12	0.45	0.10	6	0 .64	0.12	0.81	0.02
	C1	1101	25	0.58	0.13	0.54	0.18	0.94	19	0.55	0.13	0.46	0.11	6	0 .66	0.13	0.80	0.04
	C2	1191	25	0.61	0.14	0.53	0.19	0.94	19	0.58	0.13	0.44	0.12	6	0 .71	0.10	0.82	0.03
	C3	1252	25	0.61	0.14	0.52	0.20	0.94	19	0.58	0.13	0.44	0.13	6	0 .73	0.09	0.80	0.04
	D1	1030	25	0.66	0.14	0.55	0.14	0.93	19	0.66	0.14	0.48	0.08	6	0 .69	0.15	0.77	0.01
	D2	969	25	0.62	0.13	0.57	0.14	0.94	19	0.61	0.13	0.51	0.08	6	0 .66	0.16	0.78	0.02
	D3	1122	25	0.63	0.14	0.49	0.17	0.91	19	0.62	0.15	0.42	0.13	6	0 .66	0.14	0.71	0.01

* Reading aloud fluency item.

Table 4

Raw Score Summary Statistics by Modality, Grade, and Field Test Form[1, 2]

				Form 1				Form 2				Form 3			
Modality	Grade	Test Level	Max. Score	No. of Cases	Mean	SD	Median	No. of Cases	Mean	SD	Median	No. of Cases	Mean	SD	Median
Listening	K	A	20	413	10	4	10	413	10	4	10	345	11	4	11
	1	A	20	499	13	4	14	499	13	4	14	326	14	3	15
	2	A	20	406	16	4	16	406	16	4	16	319	16	3	17
	3	A	20	524	17	3	18	524	17	3	18	437	17	3	18
	4	B	28	424	19	6	21	424	19	6	21	396	19	5	19
	5	B	28	436	19	6	21	436	19	6	21	428	20	7	22
	6	B	28	346	20	6	22	346	20	6	22	407	19	6	21
	7	C	28	326	19	7	21	326	19	7	21	395	21	6	23
	8	C	28	374	20	7	22	374	20	7	22	443	22	6	24
	9	C	28	334	20	7	22	334	20	7	22	405	19	7	22
	10	D	28	303	21	6	23	303	21	6	23	483	21	6	22
	11	D	28	337	21	7	23	337	21	7	23	358	21	5	23
	12	D	28	349	22	5	23	349	22	5	23	231	23	5	23
Speaking	K	A	16	413	7	4	7	421	7	5	7	345	8	5	8
	1	A	16	499	10	4	11	434	10	4	11	326	11	4	12
	2	A	16	406	12	4	13	490	12	4	13	319	12	3	13
	3	A	16	524	13	4	14	572	13	3	14	437	13	4	14
	4	B	24	424	18	6	20	415	18	6	20	396	18	5	20
	5	B	24	436	18	6	20	330	18	7	21	428	18	6	20
	6	B	24	346	18	6	20	371	18	7	21	407	17	7	19
	7	C	24	326	16	7	19	457	17	7	19	395	16	7	18
	8	C	24	374	16	8	19	377	17	7	19	443	17	6	19
	9	C	24	334	16	7	18	327	14	8	16	405	15	7	17
	10	D	24	303	17	6	19	269	16	7	18	483	17	6	19
	11	D	24	337	19	5	20	289	18	6	20	358	18	5	19
	12	D	24	349	18	5	20	295	19	5	20	231	19	5	20

[1] Results for Reading and Writing Forms 4, 5, and 6 are combined with results for Forms 1, 2, and 3, because they are based on the same item sets.

[2] A shaded block identifies students in grades administered the same test level.

Table 4 (cont'd)
Raw Score Summary Statistics by Modality, Grade, and Field Test Form

Modality	Grade		Max. Score	Form 1 No. of Cases	Form 1 Mean	Form 1 SD	Form 1 Median	Form 2 No. of Cases	Form 2 Mean	Form 2 SD	Form 2 Median	Form 3 No. of Cases	Form 3 Mean	Form 3 SD	Form 3 Median
Reading	K	A	24	381	6	3	6	358	6	4	6	274	7	3	7
	1	A	24	483	12	4	11	405	12	5	12	315	12	4	12
	2	A	24	399	17	4	18	479	18	4	19	311	18	4	18
	3	A+Ext	34	517	25	5	26	561	25	5	26	429	25	6	26
	4	B	34	448	18	8	19	437	18	8	18	408	18	7	17
	5	B	34	445	19	8	20	360	19	8	19	440	20	8	22
	6	B	34	358	21	8	22	383	21	8	22	429	20	8	20
	7	C	34	336	18	8	17	475	15	6	15	401	18	7	17
	8	C	34	384	18	8	18	384	16	6	15	448	20	7	21
	9	C	34	376	19	8	18	331	16	7	15	400	20	8	19
	10	D	33	311	19	7	19	339	18	7	18	487	19	7	19
	11	D	33	339	20	7	20	313	22	7	23	379	20	7	21
	12	D	33	347	21	7	22	315	22	7	22	251	22	7	23
Writing	K	A	16	418	1	2	0	425	1	2	0	378	1	2	1
	1	A	16	528	6	4	6	438	5	4	5	368	6	4	5
	2	A	16	427	11	4	11	503	10	4	11	346	11	4	12
	3	A+Ext	34	545	24	7	26	580	24	8	26	455	23	8	26
	4	B	39	450	22	9	24	437	23	9	25	409	23	8	24
	5	B	39	446	23	9	26	360	23	10	24	441	24	10	27
	6	B	39	359	24	9	27	385	24	9	26	435	23	9	24
	7	C	39	339	23	9	25	475	25	9	27	401	24	9	26
	8	C	39	387	24	9	25	387	26	9	28	450	27	8	29
	9	C	39	375	23	9	24	329	24	10	26	401	25	9	26
	10	D	39	315	25	9	27	338	21	9	21	493	24	8	25
	11	D	39	359	26	9	27	316	26	9	27	380	25	7	25
	12	D	39	356	27	8	28	315	26	9	27	249	27	8	27

[1] Results for Reading and Writing Forms 4, 5, and 6 combined with results for Forms 1, 2, and 3, because they are based on the same item sets.

[2] A shaded block identifies students in grades administered the same test level.

Table 5

Listening/Speaking: Cumulative Raw Score Distributions for Field Tests by Grade

Grade K				Grade 01				Grade 02				Grade 03				Grade 04				Grade 05				Grade 06			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
0	0.6	28	93.6	1	0.1	29	80.5	0	0.2	30	58.4	0	0.1	30	39.5	0	0.2	29	15.5	1	0.1	29	18.1	0	0.1	30	19.2
1	1.7	29	95.5	2	0.2	30	85.5	1	0.3	31	69.4	2	0.2	31	49.7	2	0.2	30	16.3	2	0.3	30	19.3	2	0.3	31	20.0
2	2.4	30	97.4	3	0.5	31	89.5	3	0.6	32	79.4	4	0.4	32	61.3	3	0.3	31	17.4	3	0.5	31	19.8	4	0.4	32	21.5
3	3.1	31	98.2	4	0.7	32	93.8	4	0.8	33	88.0	5	0.7	33	73.9	4	0.6	32	18.9	4	0.8	32	21.1	5	0.6	33	22.8
4	5.0	32	98.9	5	1.4	33	96.7	5	1.8	34	94.4	6	0.8	34	86.6	5	0.6	33	20.6	5	1.3	33	22.6	6	1.1	34	24.6
5	6.2	33	99.4	6	2.3	34	98.7	6	2.2	35	98.4	7	1.2	35	95.0	6	0.8	34	22.3	6	1.7	34	24.1	7	1.9	35	26.3
6	8.7	34	100.0	7	3.2	35	99.8	7	2.6	36	100.0	8	1.8	36	100.0	7	1.1	35	24.4	7	2.3	35	25.2	8	2.3	36	28.1
7	11.2			8	3.7	36	100.0	8	2.7			9	2.5			8	1.4	36	26.1	8	2.8	36	26.7	9	2.9	37	31.1
8	14.1			9	4.5			9	3.4			10	2.9			9	2.3	37	28.4	9	3.4	37	28.4	10	3.4	38	33.5
9	16.6			10	5.4			10	3.6			11	3.1			10	2.8	38	31.4	10	4.1	38	30.1	11	3.7	39	35.1
10	19.7			11	6.4			12	4.0			12	3.6			11	3.0	39	33.9	11	4.9	39	33.5	12	4.5	40	37.4
11	23.2			12	7.1			13	4.6			13	4.0			12	3.6	40	37.2	12	6.1	40	35.6	13	5.4	41	40.4
12	27.6			13	8.3			14	5.0			14	4.2			13	4.5	41	40.5	13	6.8	41	38.2	14	6.2	42	43.6
13	31.7			14	9.7			15	5.7			15	4.7			14	4.9	42	44.9	14	7.6	42	41.6	15	6.9	43	47.4
14	35.8			15	11.6			16	6.3			16	5.4			15	5.5	43	49.1	15	8.1	43	45.1	16	7.7	44	51.4
15	40.6			16	13.2			17	7.2			17	5.9			16	6.1	44	53.4	16	8.4	44	48.7	17	8.4	45	54.9
16	45.2			17	15.7			18	7.9			18	6.9			17	6.3	45	58.9	17	9.0	45	53.5	18	8.6	46	58.8
17	49.7			18	18.8			19	8.9			19	7.5			18	6.6	46	63.8	18	9.0	46	59.7	19	9.1	47	62.8
18	54.2			19	21.5			20	10.8			20	8.3			19	7.1	47	69.7	19	9.7	47	64.8	20	10.1	48	68.5
19	59.1			20	25.4			21	12.6			21	9.5			20	8.1	48	75.7	20	10.5	48	69.6	21	10.9	49	73.6
20	64.0			21	30.6			22	14.6			22	10.4			21	8.8	49	81.7	21	11.4	49	74.8	22	11.6	50	77.8
21	69.4			22	35.3			23	17.0			23	12.2			22	9.7	50	86.9	22	11.7	50	80.4	23	12.5	51	81.4
22	74.6			23	41.4			24	19.9			24	13.6			23	10.4	51	90.6	23	12.2	51	86.3	24	13.4	52	86.8
23	78.9			24	47.4			25	25.1			25	15.2			24	11.3	52	93.7	24	13.1	52	90.4	25	13.8	53	92.2
24	82.1			25	53.9			26	30.3			26	17.5			25	11.7	53	96.6	25	14.1	53	93.7	26	14.9	54	95.8
25	85.7			26	61.5			27	36.3			27	21.4			26	12.2	54	98.7	26	15.2	54	96.8	27	16.0	55	98.5
26	88.6			27	68.3			28	42.2			28	25.0			27	13.6	55	99.8	27	16.2	55	98.8	28	16.7	56	100.0
27	91.3			28	74.5			29	50.2			29	31.3			28	14.7	56	100.0	28	17.1	56	100.0	29	17.6		

Table 5 (cont.)

Listening/Speaking: Cumulative Raw Score Distributions for Field Tests by Grade

Grade 07				Grade 08				Grade 09				Grade10				Grade 11				Grade 12			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
2	0.1	30	26.1	0	0.1	31	25.5	0	0.2	28	32.3	4	0.2	33	33.0	1	0.1	34	25.7	10	0.1	39	37.5
3	0.2	31	27.7	1	0.3	32	26.2	1	0.3	29	33.6	6	0.5	34	34.1	3	0.2	35	28.7	11	0.3	40	40.7
4	0.7	32	28.9	5	0.5	33	28.1	2	0.5	30	35.4	7	0.6	35	36.5	6	0.3	36	30.3	12	0.5	41	44.7
5	1.3	33	30.2	6	1.1	34	29.6	3	0.7	31	36.9	8	0.9	36	38.8	7	0.6	37	32.6	13	0.6	42	48.1
6	1.6	34	32.1	7	1.8	35	31.8	4	1.1	32	39.2	9	1.0	37	41.1	8	0.7	38	34.6	15	1.5	43	53.6
7	2.5	35	34.3	8	2.5	36	33.7	5	1.8	33	40.8	10	1.4	38	44.1	10	1.0	39	37.8	16	1.8	44	58.1
8	3.8	36	36.7	9	3.4	37	35.8	6	2.4	34	43.7	11	2.4	39	45.9	11	1.3	40	40.9	17	2.3	45	63.5
9	5.3	37	39.4	10	4.2	38	38.4	7	3.3	35	45.6	12	3.1	40	50.5	13	1.6	41	45.1	18	2.9	46	68.5
10	6.1	38	43.0	11	5.6	39	42.0	8	4.8	36	46.8	13	3.5	41	53.0	14	2.2	42	49.1	19	3.3	47	73.0
11	6.9	39	45.9	12	6.5	40	45.5	9	5.9	37	48.9	14	4.5	42	57.0	15	2.7	43	54.0	20	3.7	48	80.0
12	8.1	40	49.6	13	7.4	41	48.7	10	7.8	38	51.8	15	5.2	43	61.2	16	3.0	44	59.7	21	4.3	49	85.7
13	8.6	41	54.0	14	8.0	42	54.3	11	9.0	39	54.1	16	6.1	44	65.7	17	3.5	45	65.2	22	5.0	50	92.0
14	9.3	42	59.0	15	9.3	43	57.5	12	10.1	40	56.7	17	6.7	45	70.8	18	4.0	46	72.5	23	5.8	51	97.5
15	10.4	43	63.7	16	10.0	44	61.7	13	11.3	41	59.8	18	7.9	46	75.8	19	4.9	47	77.5	24	6.9	52	100.0
16	11.0	44	68.8	17	10.8	45	66.5	14	12.4	42	62.4	19	9.4	47	81.6	20	5.5	48	84.0	25	7.7		
17	12.1	45	74.6	18	11.4	46	71.8	15	13.6	43	65.5	20	10.9	48	86.8	21	6.8	49	89.5	26	8.6		
18	12.6	46	80.3	19	12.3	47	78.2	16	14.6	44	69.0	21	12.1	49	90.4	22	7.6	50	94.3	27	9.9		
19	13.3	47	84.5	20	13.5	48	83.0	17	16.1	45	73.5	22	13.8	50	95.0	23	8.6	51	97.9	28	11.5		
20	14.2	48	89.3	21	14.7	49	88.9	18	17.5	46	77.0	23	15.5	51	98.2	24	10.0	52	100.0	29	13.1		
21	15.4	49	93.0	22	15.8	50	94.1	19	18.9	47	80.6	24	17.1	52	100.0	25	10.7			30	15.0		
22	16.2	50	96.1	23	16.8	51	98.6	20	20.1	48	84.4	25	18.4			26	12.0			31	15.7		
23	17.2	51	98.6	24	18.0	52	100.0	21	21.0	49	91.1	26	20.0			27	13.3			32	17.9		
24	17.9	52	100.0	25	19.1			22	23.2	50	95.7	27	21.3			28	14.9			33	20.2		
25	18.7			26	19.9			23	24.5	51	98.7	28	22.8			29	16.2			34	22.9		
26	19.8			27	21.3			24	26.3	52	100.0	29	24.3			30	17.8			35	25.8		
27	21.2			28	22.0			25	27.5			30	26.8			31	19.9			36	28.6		
28	22.4			29	23.1			26	29.4			31	28.6			32	22.0			37	30.9		
29	24.4			30	24.1			27	30.6			32	31.0			33	23.7			38	33.8		

Table 6

Reading: Cumulative Raw Score Distributions for Field Tests by Grade

Grade K		Grade 01		Grade 02		Grade 03				Grade 04				Grade 05				Grade 06			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
0	2.9	0	0.2	1	0.1	3	0.1	31	91.4	0	0.9	28	90.9	0	0.4	28	85.4	0	0.3	28	80.4
1	8.5	1	0.7	2	0.3	4	0.1	32	96.1	1	1.6	29	93.5	1	0.9	29	89.6	1	0.4	29	85.2
2	15.3	2	1.7	3	0.6	5	0.3	33	98.3	2	2.2	30	95.9	2	1.4	30	93.2	2	0.7	30	89.7
3	22.8	3	2.4	4	1.0	6	0.3	34	100.0	3	3.1	31	97.8	3	2.4	31	95.6	3	0.9	31	92.9
4	34.6	4	4.0	5	1.3	7	0.6			4	3.8	32	98.7	4	3.2	32	98.4	4	1.4	32	95.9
5	43.6	5	6.4	6	2.0	8	1.1			5	5.1	33	99.2	5	4.4	33	99.6	5	2.1	33	98.1
6	55.1	6	9.6	7	3.4	9	1.2			6	6.2	34	100.0	6	5.7	34	100.0	6	3.2	34	100.0
7	66.0	7	14.0	8	4.6	10	1.3			7	7.9			7	7.1			7	4.7		
8	77.0	8	21.5	9	6.1	11	1.9			8	10.2			8	9.8			8	7.7		
9	84.2	9	30.0	10	7.5	12	2.6			9	13.4			9	13.3			9	10.9		
10	90.9	10	39.7	11	9.4	13	3.2			10	17.5			10	16.7			10	14.2		
11	94.4	11	49.0	12	11.9	14	4.4			11	20.8			11	19.4			11	18.5		
12	96.3	12	58.8	13	14.9	15	5.9			12	25.1			12	22.6			12	21.1		
13	97.7	13	66.4	14	19.8	16	7.0			13	29.1			13	25.5			13	23.8		
14	98.3	14	72.8	15	25.2	17	9.2			14	33.5			14	30.0			14	26.1		
15	98.8	15	78.9	16	31.7	18	11.1			15	36.5			15	32.7			15	28.7		
16	99.0	16	83.8	17	39.1	19	13.1			16	41.8			16	36.2			16	32.6		
18	99.3	17	88.3	18	50.4	20	16.5			17	46.2			17	39.5			17	35.9		
19	99.6	18	90.9	19	61.6	21	20.8			18	51.5			18	42.7			18	39.7		
20	99.8	19	93.7	20	73.2	22	25.1			19	56.3			19	47.0			19	43.8		
21	99.9	20	96.2	21	84.4	23	31.6			20	60.4			20	50.6			20	47.4		
22	100.0	21	97.6	22	92.9	24	38.1			21	64.5			21	54.7			21	50.7		
		22	99.2	23	98.3	25	47.0			22	68.8			22	60.1			22	55.0		
		23	99.7	24	100.0	26	55.3			23	73.9			23	65.1			23	59.7		
		24	100.0			27	63.2			24	77.8			24	69.0			24	63.2		
						28	71.6			25	81.5			25	72.4			25	67.0		
						29	78.5			26	85.2			26	76.8			26	71.8		
						30	84.7			27	88.2			27	80.5			27	76.5		

Table 6 (cont.)

Reading: Cumulative Raw Score Distributions for Field Tests by Grade

Grade 07				Grade 08				Grade 09				Grade 10				Grade 11		Grade 12			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
0	0.2	28	93.6	1	0.1	29	92.9	0	0.2	28	89.3	0	0.1	28	89.3	0	0.1	3	0.3	31	93.8
1	0.3	29	95.6	2	0.2	30	95.1	1	0.5	29	92.5	1	0.3	29	92.2	2	0.2	4	0.5	32	98.4
2	0.8	30	96.9	3	0.5	31	97.7	2	0.7	30	94.9	2	0.4	30	95.8	3	0.4	5	0.7	33	100.0
3	1.3	31	98.3	4	1.3	32	98.8	3	1.4	31	96.1	3	0.5	31	98.2	4	0.5	6	1.2		
4	2.5	32	99.1	5	2.5	33	99.5	4	2.8	32	98.0	4	0.8	32	99.6	5	1.5	7	2.1		
5	3.9	33	99.7	6	3.9	34	100.0	5	3.8	33	99.2	5	1.9	33	100.0	6	2.1	8	3.1		
6	6.3	34	100.0	7	6.4			6	5.8	34	100.0	6	3.3			7	2.9	9	4.5		
7	10.8			8	10.4			7	8.3			7	6.2			8	4.8	10	6.1		
8	14.4			9	13.3			8	11.9			8	9.1			9	7.0	11	8.7		
9	18.7			10	17.2			9	15.4			9	12.3			10	8.9	12	11.3		
10	23.5			11	22.0			10	19.9			10	16.5			11	11.0	13	14.3		
11	28.1			12	26.2			11	23.6			11	20.2			12	14.3	14	17.5		
12	33.2			13	31.0			12	27.8			12	24.8			13	18.3	15	21.6		
13	38.7			14	35.1			13	32.2			13	29.8			14	21.3	16	25.0		
14	43.2			15	39.1			14	37.1			14	33.5			15	25.0	17	28.3		
15	46.9			16	42.2			15	42.4			15	37.6			16	29.2	18	32.0		
16	51.8			17	46.9			16	45.0			16	41.8			17	32.5	19	36.6		
17	56.4			18	51.0			17	48.9			17	45.8			18	37.7	20	42.4		
18	61.2			19	54.9			18	53.3			18	49.9			19	42.6	21	47.0		
19	65.8			20	59.7			19	56.5			19	54.3			20	48.1	22	50.9		
20	70.0			21	63.3			20	60.3			20	58.2			21	52.8	23	55.9		
21	73.8			22	68.5			21	64.2			21	62.9			22	56.5	24	59.8		
22	77.2			23	72.3			22	67.6			22	66.7			23	60.5	25	64.0		
23	80.6			24	75.9			23	71.8			23	70.5			24	65.5	26	67.8		
24	83.4			25	79.2			24	74.7			24	74.9			25	70.6	27	74.0		
25	86.3			26	83.2			25	79.0			25	78.9			26	75.2	28	79.7		
26	88.5			27	87.4			26	82.2			26	81.7			27	79.6	29	84.7		
27	90.7			28	90.4			27	85.5			27	85.1			28	83.9	30	89.7		

Table 7

Writing: Cumulative Raw Score Distributions for Field Tests by Grade

Grade K		Grade 01		Grade 02		Grade 03				Grade 04				Grade 05				Grade 06			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
0	50.9	0	10.5	0	3.6	0	1.5	28	70.0	0	2.5	28	72.6	0	2.2	28	62.7	0	2.0	28	62.6
1	76.6	1	19.3	1	5.7	1	2.0	29	77.3	1	2.9	29	76.9	1	2.6	29	68.6	1	2.1	29	67.4
2	88.6	2	26.5	2	7.8	2	2.7	30	82.8	2	3.1	30	80.9	2	2.7	30	73.5	2	2.4	30	72.8
3	92.1	3	34.5	3	9.2	3	3.5	31	89.4	3	3.5	31	85.3	3	3.8	31	78.1	3	2.8	31	78.2
4	94.7	4	43.1	4	11.1	4	3.9	32	94.8	4	4.2	32	88.9	4	5.3	32	83.6	4	3.6	32	82.3
5	96.7	5	52.4	5	13.4	5	4.6	33	98.4	5	5.6	33	91.9	5	6.3	33	87.0	5	5.2	33	86.1
6	97.8	6	61.2	6	17.5	6	5.1	34	100.0	6	6.9	34	95.0	6	8.0	34	90.9	6	5.9	34	89.5
7	98.9	7	70.2	7	23.1	7	5.9			7	8.1	35	97.0	7	9.4	35	94.5	7	7.0	35	92.5
8	99.3	8	78.6	8	29.4	8	6.3			8	9.3	36	98.5	8	10.7	36	97.4	8	7.7	36	95.3
9	99.4	9	85.2	9	37.0	9	7.1			9	11.1	37	99.3	9	11.8	37	98.6	9	9.6	37	97.8
10	99.5	10	89.4	10	43.9	10	8.0			10	12.1	38	99.8	10	13.2	38	99.4	10	10.9	38	99.2
11	99.8	11	92.4	11	52.7	11	8.9			11	13.3	39	100.0	11	15.0	39	100.0	11	13.1	39	100.0
13	100.0	12	94.8	12	62.8	12	10.3			12	15.0			12	17.0			12	15.2		
		13	96.9	13	72.5	13	11.3			13	16.7			13	18.4			13	16.6		
		14	98.4	14	81.4	14	12.2			14	18.4			14	20.6			14	18.2		
		15	99.3	15	90.0	15	14.4			15	20.5			15	22.1			15	20.0		
		16	100.0	16	100.0	16	16.1			16	23.8			16	23.7			16	21.7		
						17	18.7			17	26.3			17	25.7			17	24.1		
						18	22.0			18	28.6			18	27.3			18	26.0		
						19	24.1			19	31.9			19	29.3			19	29.9		
						20	26.4			20	35.4			20	31.4			20	32.8		
						21	29.8			21	39.4			21	34.1			21	35.9		
						22	33.4			22	43.4			22	38.2			22	38.8		
						23	38.8			23	47.4			23	41.5			23	41.8		
						24	43.8			24	52.2			24	45.8			24	46.1		
						25	49.5			25	57.3			25	49.6			25	49.5		
						26	55.4			26	61.9			26	53.7			26	53.3		
						27	62.8			27	67.4			27	57.8			27	57.7		

Table 7 (cont.)

Writing: Cumulative Raw Score Distributions for Field Tests by Grade

Grade 07				Grade 08				Grade 09				Grade 10				Grade 11				Grade 12			
RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt	RS	cum. pcnt
0	0.4	28	61.2	0	0.3	28	55.3	0	0.6	28	62.1	0	1.9	30	76.0	0	2.5	29	63.0	0	2.2	30	64.5
1	0.7	29	67.3	1	0.6	29	60.4	1	0.8	29	65.3	3	2.5	31	80.8	1	2.7	30	68.6	2	2.3	31	69.3
2	1.2	30	72.7	2	0.9	30	64.5	2	1.1	30	70.0	4	3.2	32	84.4	2	2.9	31	73.5	4	2.4	32	73.6
3	2.1	31	76.5	3	1.3	31	69.6	3	1.7	31	73.4	5	3.7	33	88.0	3	3.0	32	78.6	5	2.6	33	78.8
4	3.0	32	81.6	4	2.5	32	74.0	4	2.3	32	76.8	6	4.5	34	91.6	4	3.3	33	83.9	6	2.7	34	83.3
5	4.3	33	85.3	5	3.4	33	80.3	5	2.9	33	80.9	7	5.5	35	94.2	5	3.6	34	87.6	7	2.8	35	87.7
6	5.1	34	88.9	6	4.0	34	85.0	6	3.9	34	85.2	8	6.4	36	96.3	6	3.9	35	91.1	8	3.2	36	92.0
7	6.7	35	92.8	7	4.9	35	90.0	7	5.0	35	89.2	9	7.7	37	97.9	7	4.5	36	94.8	9	3.3	37	95.5
8	8.1	36	96.0	8	5.9	36	93.6	8	7.1	36	93.8	10	9.2	38	99.4	9	5.2	37	96.7	10	3.8	38	98.6
9	9.1	37	97.7	9	6.6	37	96.0	9	8.8	37	96.4	11	10.7	39	100.0	10	6.2	38	98.9	11	4.2	39	100.0
10	10.9	38	99.2	10	7.8	38	98.6	10	10.3	38	98.7	12	13.1			11	6.9	39	100.0	12	5.0		
11	12.8	39	100.0	11	9.1	39	100.0	11	12.6	39	100.0	13	15.4			12	8.1			13	6.5		
12	14.1			12	10.6			12	14.3			14	17.9			13	9.0			14	7.7		
13	15.6			13	12.0			13	16.4			15	19.7			14	10.8			15	9.5		
14	17.0			14	13.7			14	19.0			16	22.5			15	12.4			16	10.9		
15	18.4			15	15.5			15	21.1			17	25.6			16	14.1			17	13.3		
16	20.4			16	17.6			16	23.6			18	28.7			17	16.3			18	14.6		
17	22.9			17	19.4			17	26.3			19	32.2			18	19.1			19	17.1		
18	24.9			18	21.8			18	29.1			20	35.6			19	22.3			20	19.5		
19	27.4			19	24.4			19	32.1			21	38.9			20	24.8			21	22.4		
20	30.3			20	27.2			20	34.8			22	41.6			21	27.7			22	26.3		
21	32.9			21	30.5			21	37.7			23	44.9			22	31.0			23	30.9		
22	35.7			22	32.8			22	40.9			24	49.2			23	35.6			24	35.3		
23	39.3			23	35.8			23	43.9			25	53.1			24	40.6			25	40.0		
24	43.0			24	39.0			24	47.2			26	57.1			25	45.1			26	44.6		
25	47.5			25	42.2			25	50.8			27	62.1			26	50.4			27	50.1		
26	52.1			26	46.1			26	54.4			28	66.8			27	54.3			28	54.6		
27	55.8			27	50.7			27	58.5			29	70.6			28	58.8			29	59.7		

Table 8

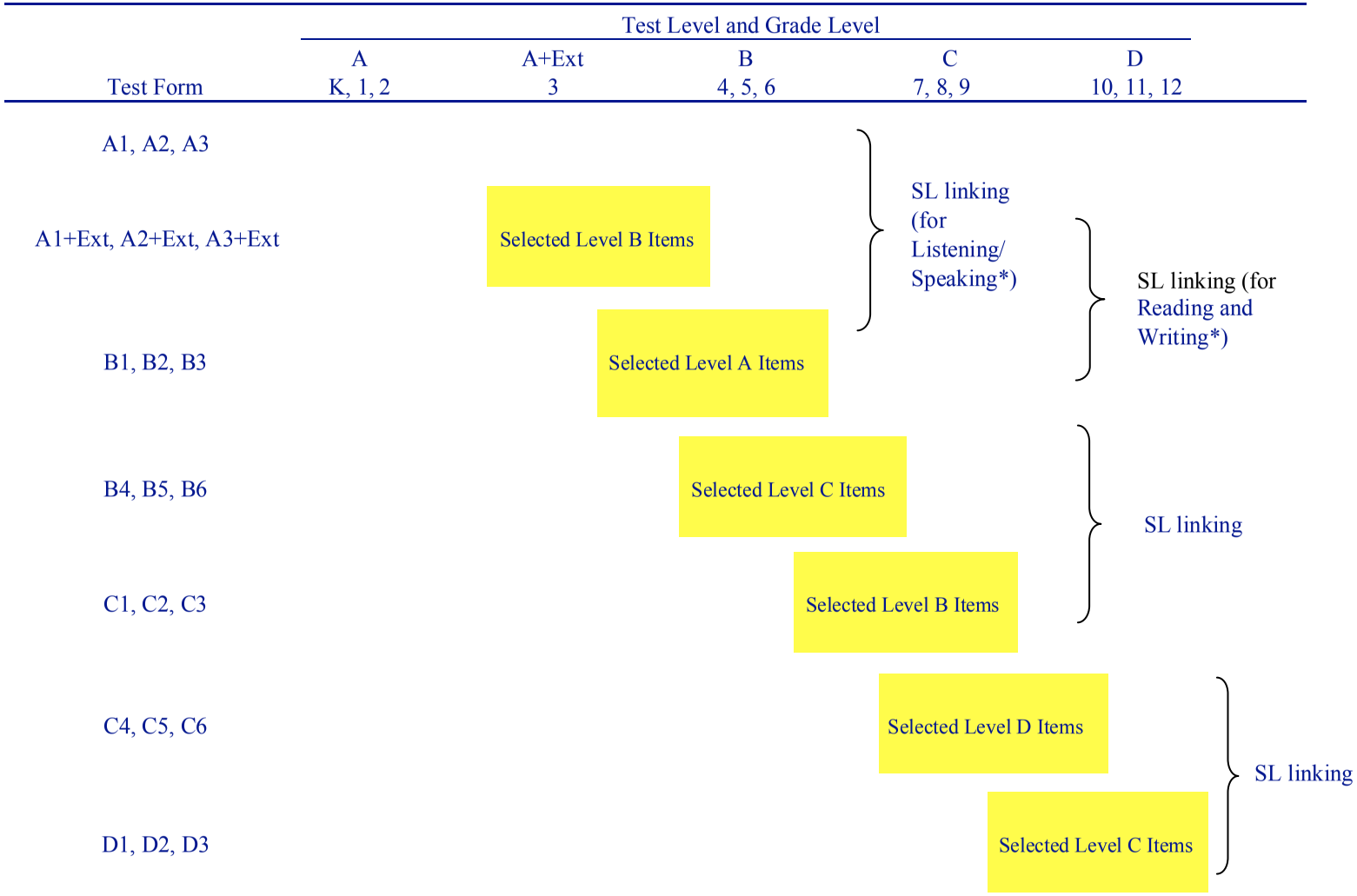
Calibration Design for the Field Test and Vertical Linking Items

			Sets of Items Calibrated Together	
Grades	Calibration Run	Test Level	On-Level Items	Vertical Linking Items
K, 1, 2, 3	1	A	A1, A2, A3, A1+Ext, A2+Ext, A3+Ext[1]	Level B
4, 5, 6	2	B	B1, B2, B3 B4, B5, B6	Level A Level C
7, 8, 9	3	C	C1, C2, C3 C4, C5, C6	Level B Level D
10, 11, 12	4	D	D1, D2, D3	Level C

[1]These forms were included in the calibration runs for Reading and Writing.

Table 9

Vertical Linking Design



*Items from A+Ext forms were used for linking Reading and for Writing; items from A forms were used for linking Listening/Speaking

Table 10

Scale Score Summary Statistics by Modality and Grade Based on Field Test Forms

Modality	Grade	Test Level	N	Mean	SD	Median	% at Loss
Listening/Speaking	K	A	1179	617	45	626	6%
	1	A	1259	651	36	655	1%
	2	A	1215	672	39	675	2%
	3	A	1533	686	39	690	1%
	4	B	1235	698	38	703	1%
	5	B	1194	700	45	707	2%
	6	B	1124	701	45	704	2%
	7	C	1178	694	44	707	5%
	8	C	1194	698	43	709	3%
	9	C	1066	687	48	700	6%
	10	D	1055	703	38	713	2%
	11	D	984	712	32	719	1%
	12	D	875	715	28	719	0%
Reading	K	A	1013	404	79	345	55%
	1	A	1203	541	99	561	10%
	2	A	1189	646	79	655	2%
	3	A+Ext	1507	677	65	687	1%
	4	B	1293	688	46	697	10%
	5	B	1245	694	48	706	10%
	6	B	1170	700	49	710	8%
	7	C	1212	711	53	724	14%
	8	C	1216	720	49	731	10%
	9	C	1107	718	52	727	11%
	10	D	1137	733	49	744	9%
	11	D	1031	746	41	751	5%
	12	D	913	752	38	755	3%
Writing	K	A	1221	550	40	515	51%
	1	A	1334	619	47	630	10%
	2	A	1276	665	45	670	4%
	3	A+Ext	1580	683	38	688	2%
	4	B	1296	693	43	699	4%
	5	B	1247	697	48	707	5%
	6	B	1179	700	47	707	4%
	7	C	1215	712	52	723	4%
	8	C	1224	721	52	727	3%
	9	C	1105	713	54	719	3%
	10	D	1146	709	44	710	3%
	11	D	1055	718	43	717	3%
	12	D	920	724	43	721	2%

Table 11

Listening/Speaking: Cumulative Scale Score Distributions for Field Tests by Grade

Grade K						Grade 01						Grade 02						Grade 03					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
495	5.6	618	41.6	675	96.3	495	1.1	619	12.5	679	84.1	495	1.5	632	9.2	705	88.0	495	0.5	622	5.5	686	45.9
524	6.6	619	42.8	679	96.8	524	1.4	620	12.9	681	85.5	530	1.8	633	9.5	712	90.0	524	0.6	623	5.7	687	49.7
530	7.2	620	44.0	681	97.4	530	1.7	622	13.7	682	87.1	532	2.0	635	10.5	715	92.0	530	0.8	624	6.1	690	53.4
532	8.2	622	46.4	682	97.9	532	1.8	623	14.7	686	88.4	554	2.2	637	10.9	718	94.4	532	0.8	626	6.3	693	56.9
549	9.0	623	48.4	686	98.1	549	2.1	624	15.0	687	89.5	564	2.4	639	12.4	732	95.4	549	0.9	627	6.7	695	61.3
551	9.8	624	49.6	687	98.2	551	2.3	626	16.0	690	90.9	568	2.7	640	12.9	735	96.5	551	1.0	628	6.9	700	64.8
554	10.3	626	50.9	690	98.7	554	2.8	627	17.5	693	92.4	574	3.1	643	14.4	738	98.4	563	1.3	630	7.2	703	68.9
563	11.1	627	52.7	695	98.9	564	3.0	628	17.9	695	93.8	583	3.4	644	15.1	755	100.0	564	1.4	631	7.5	705	73.9
564	12.0	628	54.0	700	99.3	568	3.4	630	19.2	700	94.7	586	3.6	647	16.6			568	1.6	632	7.6	712	77.3
568	12.9	630	55.6	703	99.4	574	4.0	631	20.3	703	95.6	597	3.7	648	17.5			574	2.0	635	8.3	715	82.1
574	14.7	631	56.8	712	99.7	578	4.2	632	21.0	705	96.7	602	3.9	651	21.6			578	2.1	637	8.6	718	86.6
578	15.9	632	58.0	715	99.9	583	4.8	633	22.2	712	97.7	603	4.1	655	26.4			583	2.2	639	9.3	732	89.5
583	18.1	633	60.3	718	100.0	586	5.1	635	24.7	715	98.3	604	4.4	659	27.7			586	2.6	640	9.7	735	92.2
586	18.8	635	63.5			590	5.3	637	26.8	718	98.7	609	4.8	660	31.6			593	2.9	643	10.4	738	95.0
590	19.6	637	65.4			591	5.8	639	29.2	732	99.2	611	4.9	664	37.8			596	3.2	644	10.7	755	100.0
591	21.3	639	69.1			593	6.1	640	31.3	735	99.4	614	5.4	669	44.2			597	3.3	647	11.6		
593	22.1	640	71.1			596	6.3	643	34.6	738	99.8	616	5.5	674	46.8			599	3.5	648	12.5		
596	23.2	643	74.1			597	6.7	644	36.8	755	100.0	618	5.8	675	51.9			602	3.7	651	14.1		
597	24.8	644	75.7			599	7.0	647	40.0			619	6.3	679	55.3			603	3.8	655	16.0		
599	25.8	647	78.3			602	7.7	648	42.5			620	6.6	681	58.4			604	3.9	659	16.8		
602	27.1	648	79.8			603	7.9	651	49.2			622	6.7	682	62.1			607	4.0	660	18.3		
603	28.8	651	83.4			604	7.9	655	55.8			623	6.9	686	65.8			609	4.1	664	22.2		
604	30.6	655	86.6			607	8.2	659	57.3			624	7.0	687	69.4			611	4.2	669	27.1		
607	31.7	659	87.3			609	9.3	660	63.1			626	7.2	690	72.2			614	4.6	674	28.9		
609	34.4	660	89.3			611	9.8	664	69.9			627	7.7	693	74.9			616	4.7	675	33.5		
611	36.2	664	92.0			614	11.0	669	76.6			628	8.0	695	79.4			618	4.8	679	36.5		
614	38.8	669	94.3			616	11.4	674	78.6			630	8.1	700	82.1			619	5.0	681	39.5		
616	39.8	674	94.8			618	12.0	675	82.1			631	8.6	703	85.2			620	5.1	682	42.1		

Table 11 (cont.)

Listening/Speaking: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 04								Grade 05								Grade 06							
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
560	1.1	647	9.1	680	24.1	720	73.2	560	2.3	646	10.9	679	25.2	718	65.9	560	1.9	647	10.9	680	26.5	720	66.0
573	1.2	648	9.4	681	26.1	721	75.1	574	2.7	647	11.1	680	25.8	720	67.5	573	2.0	648	11.2	681	27.8	721	67.8
574	1.4	649	10.0	683	27.0	722	77.6	579	2.8	648	11.6	681	26.8	721	68.9	574	2.2	649	11.7	683	28.9	722	70.3
590	1.6	651	10.4	684	28.8	724	80.4	590	2.9	649	12.0	683	27.4	722	71.0	579	2.3	651	12.4	684	30.2	724	72.2
591	2.2	652	11.0	686	29.9	726	82.8	591	3.4	651	12.5	684	28.2	724	73.0	590	2.5	652	12.5	686	32.3	726	73.2
595	2.3	653	11.1	687	30.9	727	84.0	600	3.6	652	12.9	686	29.6	726	74.6	595	2.9	653	13.0	687	32.7	727	74.6
600	2.5	654	11.4	688	31.7	730	85.7	602	3.9	653	13.1	687	30.4	727	76.4	600	3.3	654	13.3	688	33.0	730	76.0
602	2.8	655	11.6	689	34.4	732	88.4	605	4.1	654	13.6	688	31.8	730	78.2	602	3.4	655	13.8	689	34.7	732	79.3
607	2.8	656	11.8	691	35.5	736	90.0	607	4.4	655	13.7	689	32.8	732	82.2	609	3.6	656	14.3	691	35.2	736	80.3
609	3.0	658	12.6	692	37.1	739	91.7	609	4.8	656	14.4	691	33.4	736	84.1	612	3.7	658	14.6	692	37.2	739	83.8
613	3.4	659	12.9	694	39.4	744	93.0	612	4.9	658	15.2	692	35.1	739	87.8	613	4.0	659	14.9	694	39.1	744	85.2
615	3.5	660	13.4	695	41.0	746	93.7	613	5.4	659	15.2	694	37.6	744	88.6	615	4.3	660	15.6	695	39.9	746	86.8
618	3.9	661	13.8	697	43.3	748	94.7	615	5.8	660	15.7	695	38.4	746	90.4	618	4.8	661	15.7	697	42.4	748	89.3
620	4.0	662	14.6	698	44.6	753	95.6	618	6.2	661	16.2	697	40.6	748	91.9	620	5.1	662	16.6	698	44.0	753	90.9
623	4.5	663	14.8	700	47.3	756	96.6	620	6.4	662	17.0	698	41.9	753	92.8	623	5.6	663	17.0	700	46.3	756	92.2
624	4.8	665	15.1	701	49.1	760	97.7	623	7.3	663	17.3	700	44.4	756	93.7	624	6.0	665	17.3	701	47.9	760	93.5
627	5.3	666	15.5	703	50.9	765	98.4	624	7.4	665	17.8	701	44.6	760	95.0	627	6.5	666	18.0	703	49.0	765	95.2
630	5.4	667	16.4	704	51.9	769	98.7	627	7.6	666	18.3	703	45.4	765	96.1	628	6.7	667	18.4	704	50.4	769	95.8
631	5.7	668	16.8	705	53.4	781	99.3	628	8.0	667	18.9	704	47.2	769	96.8	631	7.5	668	18.7	705	51.5	781	97.5
633	5.8	669	17.9	707	57.2	785	99.5	631	8.2	668	19.3	705	48.4	781	97.7	633	7.7	669	20.1	707	53.6	785	98.0
634	6.3	670	18.2	708	58.7	791	99.8	633	8.5	669	19.9	707	51.6	785	98.2	634	8.2	670	20.6	708	55.1	791	98.5
636	6.5	672	18.8	710	60.4	805	100.0	634	8.9	670	20.7	708	52.8	791	98.8	636	8.4	672	21.3	710	56.4	805	100.0
639	6.8	673	19.4	711	62.1			636	9.0	672	21.0	710	54.8	805	100.0	637	8.6	673	21.6	711	58.1		
640	6.9	674	20.2	712	64.6			637	9.0	673	21.9	711	56.5			639	8.7	674	23.1	712	58.9		
642	7.7	675	20.5	714	66.3			639	9.5	674	22.8	712	58.0			642	9.1	675	23.6	714	59.6		
643	7.9	676	21.9	715	68.2			642	9.9	675	23.3	714	61.1			643	9.4	676	24.6	715	61.0		
644	8.4	678	22.2	716	69.6			643	10.1	676	23.8	715	63.1			644	9.9	678	25.3	716	62.7		
646	8.6	679	23.6	718	71.1			644	10.6	678	24.1	716	64.2			646	10.5	679	26.2	718	64.5		

Table 11 (cont.)

Listening/Speaking: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 07								Grade 08								Grade 09							
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
565	5.3	652	13.9	683	26.9	722	78.4	565	3.4	653	13.5	684	25.8	725	71.8	565	5.9	650	19.2	681	35.9	719	71.8
580	5.7	653	14.2	684	28.1	725	80.3	580	3.5	654	13.9	685	25.9	726	75.9	580	6.5	652	19.3	683	36.5	722	75.7
583	5.8	654	14.5	685	28.4	726	83.3	583	3.9	655	14.3	686	27.1	729	79.2	583	7.1	653	20.1	684	37.4	725	77.0
586	6.1	655	14.7	686	29.5	729	85.5	586	4.2	656	14.7	687	27.2	730	80.5	586	7.8	654	20.4	685	38.2	726	79.5
597	6.3	656	15.4	687	29.8	730	87.1	597	4.7	657	15.1	689	28.9	733	83.0	597	8.1	655	20.6	686	39.6	729	81.9
599	6.5	657	15.8	689	31.0	733	89.3	599	5.2	658	15.3	690	29.4	734	86.4	599	8.4	656	21.0	687	40.1	730	83.3
601	6.9	658	15.9	690	31.6	734	91.9	601	5.6	659	15.8	691	30.8	738	90.4	601	9.0	657	21.8	689	41.7	733	84.4
609	7.4	659	16.2	691	33.0	738	94.0	609	6.2	660	16.2	692	31.2	739	91.6	609	9.6	658	22.7	690	42.9	734	88.4
610	7.6	660	16.8	692	33.7	739	95.0	610	6.4	661	16.3	694	32.4	743	94.1	610	9.9	659	23.2	691	44.2	738	92.3
612	8.1	661	17.0	694	35.0	743	96.1	612	6.5	662	16.8	695	33.1	744	95.6	612	10.1	660	23.5	692	44.7	739	93.3
617	8.3	662	17.2	695	35.7	744	96.6	617	6.7	663	16.8	696	34.3	745	97.2	617	10.4	661	23.7	694	46.0	743	95.7
619	8.4	663	17.6	696	37.9	745	98.0	619	7.2	664	17.6	697	34.8	749	98.6	619	10.8	662	24.5	695	46.2	744	96.9
620	8.6	664	17.8	697	38.5	749	98.6	620	7.4	665	18.0	699	36.7	750	99.2	620	11.3	663	25.4	696	47.7	745	97.7
624	9.0	665	17.9	699	40.5	750	99.1	626	8.0	666	18.1	700	37.7	752	99.7	624	11.8	664	25.8	697	48.5	749	98.7
626	9.3	666	18.4	700	41.3	752	99.8	630	8.4	667	18.5	701	39.7	757	100.0	626	12.4	665	26.3	699	49.6	750	99.1
630	9.8	667	18.5	701	44.1	757	100.0	631	9.0	668	19.2	702	41.0			630	12.9	666	26.5	700	50.6	752	99.4
631	10.2	668	19.0	702	45.0			632	9.3	670	19.3	704	43.1			631	13.1	667	27.2	701	52.6	757	100.0
632	10.4	670	19.5	704	47.5			635	9.4	671	20.5	705	43.9			632	13.6	668	28.4	702	53.0		
635	10.6	671	20.2	705	48.1			636	9.7	672	21.0	707	47.2			635	13.9	670	28.8	704	54.9		
636	10.7	672	21.0	707	53.1			637	10.0	673	21.3	709	50.5			636	14.3	671	29.4	705	55.6		
637	11.0	673	21.2	709	56.4			639	10.2	674	21.5	710	52.6			637	14.6	672	29.7	707	58.4		
639	11.4	674	21.5	710	57.5			641	10.8	675	21.8	712	55.3			639	15.0	673	30.6	709	60.3		
641	12.1	675	21.9	712	61.0			643	11.1	676	22.4	713	55.9			641	16.1	674	31.0	710	61.3		
645	12.6	676	22.9	713	62.2			645	11.4	678	22.7	715	58.8			643	16.3	675	31.1	712	63.3		
647	12.9	678	23.7	715	65.7			648	11.9	679	23.4	716	59.9			645	17.5	676	32.6	713	64.4		
648	13.1	679	24.9	716	66.9			649	12.3	680	23.7	718	63.1			647	17.7	678	33.0	715	66.1		
649	13.3	680	25.6	718	70.5			650	12.7	681	24.8	719	64.2			648	18.1	679	34.3	716	68.0		
650	13.4	681	26.7	719	72.1			652	13.1	683	25.1	722	70.2			649	18.9	680	34.9	718	69.8		

Table 11 (cont.)

Listening/Speaking: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 10						Grade 11						Grade 12					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
580	2.5	663	14.7	708	45.6	580	1.2	672	11.1	716	46.1	580	0.2	679	11.0	725	60.6
583	2.6	666	16.0	710	46.5	583	1.3	673	11.3	717	47.7	583	0.5	681	11.8	726	61.7
596	2.7	667	16.4	711	49.4	611	1.4	674	11.6	719	51.6	615	0.6	682	12.1	728	64.2
600	2.8	669	17.3	713	51.2	618	1.7	675	12.1	720	53.0	622	0.8	684	14.1	729	67.1
605	3.2	670	17.8	714	52.3	622	1.9	677	12.7	722	57.0	630	0.9	686	14.3	732	70.2
611	3.6	672	18.8	716	55.6	625	2.0	678	14.1	723	57.9	632	1.1	687	15.2	733	71.5
615	3.7	673	19.1	717	56.3	627	2.4	679	14.5	725	61.8	634	1.6	689	17.3	736	77.5
618	3.9	674	19.2	719	60.0	632	2.6	681	15.9	726	63.5	638	1.8	692	19.1	740	83.9
622	4.2	675	20.8	720	60.7	634	2.7	682	16.1	728	66.3	644	2.3	694	20.7	745	89.4
625	4.5	677	21.2	722	64.5	637	2.9	684	17.2	729	70.8	647	2.4	695	21.8	749	92.0
627	4.7	678	21.8	723	65.1	638	3.0	686	18.0	732	74.5	648	2.5	697	25.4	750	95.7
630	5.1	679	22.3	725	69.1	640	3.3	687	19.1	733	76.1	649	3.0	700	27.7	755	97.5
632	5.6	681	23.4	726	69.7	642	3.6	689	21.2	736	82.4	651	3.2	702	29.0	756	98.4
637	5.8	682	23.5	728	71.7	644	3.8	692	23.4	740	87.6	653	3.5	703	29.9	757	99.2
638	6.1	684	26.1	729	74.0	647	4.1	694	24.6	745	92.9	655	3.8	705	31.8	761	100.0
640	6.3	686	26.8	732	78.2	648	4.3	695	24.9	749	94.3	656	3.9	706	32.7		
642	6.7	687	28.2	733	80.0	651	4.7	697	28.0	750	96.5	660	4.3	707	33.8		
644	7.2	689	30.3	736	85.2	652	5.1	700	29.8	755	97.9	663	5.0	708	36.1		
647	7.8	692	32.5	740	90.0	653	5.3	702	31.7	756	98.3	666	5.6	710	37.7		
648	8.2	694	33.4	745	94.3	655	5.6	703	32.2	757	99.2	667	6.1	711	39.9		
649	8.4	695	33.8	749	95.0	656	5.8	705	33.3	761	100.0	669	6.6	713	43.1		
651	8.9	697	35.8	750	97.7	659	6.0	706	33.7			670	7.1	714	43.9		
652	9.7	700	38.0	755	98.2	660	7.0	707	34.7			672	7.7	716	46.2		
653	9.9	702	40.0	756	98.9	663	8.0	708	37.0			673	7.8	717	47.0		
655	10.5	703	40.8	757	99.9	666	8.5	710	37.7			674	8.2	719	51.0		
656	11.6	705	43.0	761	100.0	667	8.6	711	40.1			675	8.8	720	52.1		
659	12.5	706	43.4			669	9.9	713	42.9			677	9.3	722	55.4		
660	13.1	707	44.2			670	10.3	714	43.6			678	10.3	723	56.9		

Table 12

Reading: Cumulative Scale Score Distributions for Field Tests by Grade

Grade K				Grade 01				Grade 02				Grade 03					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
345	55.1	649	99.2	345	9.6	628	83.8	345	2.0	638	34.2	345	1.2	637	16.5	737	93.1
377	57.6	655	99.3	377	10.7	638	85.3	377	2.7	641	39.1	377	1.3	646	19.4	744	94.8
400	61.9	668	99.4	400	12.3	641	88.3	400	2.9	649	42.5	426	1.7	647	20.8	748	96.1
418	66.0	671	99.6	418	14.0	649	88.9	418	3.4	654	45.6	451	1.9	655	22.6	759	97.2
426	69.8	675	99.8	426	16.1	654	89.5	441	3.7	655	50.4	467	2.3	656	25.1	762	98.0
441	73.0	691	99.9	441	19.4	655	90.9	459	4.6	661	54.8	485	2.5	664	29.7	773	98.3
459	77.0	722	100.0	459	21.5	661	92.0	467	5.0	668	57.5	499	2.6	665	31.6	800	100.0
467	79.1			467	23.2	668	92.8	475	5.5	671	61.6	504	2.7	672	33.5		
475	81.8			475	26.8	671	93.7	494	6.1	675	66.2	515	3.0	673	36.7		
494	84.2			494	30.0	675	94.6	504	6.4	683	69.3	530	3.2	674	38.1		
504	86.6			504	32.3	683	95.3	506	6.7	687	73.2	536	3.7	679	40.9		
506	88.5			506	36.7	687	96.2	525	7.5	691	77.9	543	4.1	681	44.3		
525	90.9			525	39.7	691	96.5	535	8.0	701	81.0	557	4.4	683	47.0		
535	92.1			535	44.1	701	96.9	537	8.6	704	84.4	564	4.7	687	50.0		
537	93.4			537	46.1	704	97.6	553	9.4	712	88.8	567	5.4	689	53.2		
553	94.4			553	49.0	712	98.4	561	10.3	722	90.4	579	5.9	691	55.3		
561	94.9			561	53.2	722	98.8	564	10.8	724	92.9	584	6.0	694	58.1		
564	95.6			564	55.9	724	99.2	576	11.9	740	95.9	587	6.7	698	60.7		
576	96.3			576	58.8	740	99.4	583	12.8	751	96.8	595	7.0	699	63.2		
583	96.8			583	61.5	751	99.6	586	13.5	754	98.3	601	7.8	701	65.8		
586	97.5			586	63.8	754	99.7	593	14.9	790	100.0	603	8.4	706	69.1		
593	97.7			593	66.4	790	100.0	600	16.5			607	9.2	707	71.6		
600	97.9			600	68.2			602	18.2			614	9.8	708	74.1		
602	98.2			602	70.6			606	19.8			616	10.4	714	78.5		
606	98.3			606	72.8			615	21.9			618	11.1	715	80.6		
615	98.5			615	75.5			616	23.5			625	11.9	723	84.7		
617	98.8			616	76.9			617	25.2			627	13.1	724	87.1		
628	99.0			617	78.9			628	31.7			636	14.1	732	89.3		

Table 12 (cont.)

Reading: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 04						Grade 05						Grade 06					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
590	10.2	701	54.8	764	98.5	590	9.8	701	45.6	764	97.5	590	7.7	701	42.1	764	95.4
618	11.2	702	56.3	769	98.7	618	10.8	702	47.0	769	98.4	618	8.7	702	43.8	769	95.9
620	12.6	704	59.2	773	99.0	620	11.7	704	49.6	773	98.8	620	9.9	704	46.5	773	97.0
625	13.4	706	60.4	786	99.2	625	13.3	706	50.6	781	99.3	625	10.9	706	47.4	781	97.5
637	14.9	708	63.5	810	100.0	637	14.6	708	52.9	786	99.6	637	11.6	708	49.1	786	98.1
641	17.5	710	64.5			641	16.7	710	54.7	810	100.0	641	14.2	710	50.7	810	100.0
649	18.2	711	66.0			649	17.7	711	56.6			649	15.0	711	52.4		
653	19.6	712	67.4			653	18.0	712	57.8			653	17.0	712	53.2		
654	20.8	714	68.8			654	19.4	714	60.1			654	18.5	714	55.0		
658	22.1	715	71.9			658	20.6	715	62.7			658	19.4	715	58.3		
662	23.7	718	73.9			662	21.6	718	65.1			662	20.5	718	59.7		
664	25.1	719	76.7			664	22.6	719	67.6			664	21.1	719	62.0		
666	26.5	722	80.4			666	23.7	722	70.8			666	21.9	722	65.6		
670	27.6	726	83.8			670	24.5	726	75.1			670	23.2	726	69.8		
671	29.1	729	84.7			671	25.5	729	75.8			671	23.8	729	71.7		
673	30.2	731	87.4			673	27.4	731	78.9			673	24.4	731	75.3		
676	32.2	734	88.2			676	28.8	734	80.5			676	25.2	734	76.5		
677	33.5	735	90.1			677	30.0	735	83.4			677	26.1	735	78.8		
680	34.3	738	91.1			680	30.9	738	84.3			680	26.7	738	80.3		
682	35.5	740	92.5			682	32.0	740	87.7			682	27.4	740	83.2		
683	36.5	743	93.3			683	32.7	743	88.8			683	28.7	743	85.3		
686	38.3	745	94.3			686	34.4	745	90.6			686	29.9	745	87.3		
688	41.8	746	95.6			688	36.2	746	91.9			688	32.6	746	88.4		
691	42.9	750	96.7			691	37.3	750	92.5			691	33.4	750	89.7		
693	46.2	752	97.0			693	39.5	752	93.8			693	35.9	752	91.0		
695	48.2	754	97.5			695	40.8	754	95.0			695	36.5	754	92.0		
697	51.5	758	98.1			697	42.7	758	96.1			697	39.7	758	93.3		
700	53.7	759	98.3			700	44.4	759	96.6			700	41.0	759	94.3		

Table 12 (cont.)

Reading: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 07						Grade 08						Grade 09					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
600	13.7	733	60.6	781	97.3	600	9.6	733	52.1	781	96.8	600	11.2	733	54.3	781	95.6
647	14.9	734	61.7	783	97.9	647	10.5	734	53.5	783	97.3	647	12.1	734	55.2	783	96.3
659	16.8	735	64.3	788	98.3	659	12.1	735	55.3	788	97.6	659	13.7	735	56.4	788	96.9
670	18.1	737	66.7	793	98.8	670	13.0	737	57.9	793	97.7	670	15.3	737	58.6	793	97.5
677	19.2	739	69.1	797	99.2	677	14.8	739	59.3	797	98.4	677	16.5	739	59.9	797	98.4
681	21.1	740	71.3	802	99.3	681	16.0	740	62.9	802	99.1	681	18.2	740	62.3	802	99.0
687	22.6	742	72.9	815	100.0	687	17.7	742	64.1	815	100.0	687	19.5	742	63.6	815	100.0
691	23.8	743	75.0			691	18.8	743	66.7			691	21.2	743	66.6		
694	26.2	746	77.4			694	19.9	746	69.3			694	22.9	746	68.9		
696	27.6	747	78.9			696	20.7	747	71.2			696	24.4	747	69.6		
700	29.1	749	79.9			700	22.1	749	71.8			700	25.8	749	70.9		
703	31.8	750	81.9			703	24.2	750	74.4			703	26.9	750	74.1		
704	33.8	752	82.4			704	26.2	752	75.6			704	27.6	752	75.2		
706	35.0	754	83.2			706	27.7	754	77.5			706	29.7	754	76.6		
709	36.5	755	84.4			709	28.9	755	79.0			709	31.9	755	77.5		
711	38.5	756	85.0			711	30.9	756	80.8			711	33.2	756	78.9		
712	39.7	758	86.0			712	32.3	758	82.7			712	34.4	758	80.0		
714	40.8	759	87.2			714	33.4	759	84.0			714	36.7	759	81.3		
716	41.9	760	88.1			716	34.6	760	84.6			716	38.6	760	82.1		
717	43.6	762	89.8			717	35.9	762	86.2			717	39.8	762	84.3		
718	44.9	764	91.3			718	36.8	764	87.5			718	41.1	764	86.7		
720	46.1	767	92.2			720	37.8	767	89.2			720	41.9	767	88.4		
722	49.4	768	93.1			722	40.8	768	90.7			722	45.0	768	89.3		
724	50.1	769	94.2			724	42.4	769	91.8			724	46.1	769	90.9		
726	52.5	773	94.6			726	45.6	773	92.4			726	48.6	773	92.4		
727	53.9	774	95.5			727	46.7	774	92.9			727	50.3	774	93.1		
730	55.2	775	96.5			730	48.2	775	93.9			730	51.5	775	93.8		
731	59.2	780	96.9			731	50.5	780	96.1			731	52.8	780	94.3		

Table 12 (cont.)

Reading: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 10						Grade 11						Grade 12					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
605	9.1	746	54.4	791	95.2	605	4.8	746	43.1	791	93.3	605	3.1	746	37.2	791	89.3
673	10.4	747	55.9	792	96.4	673	5.3	747	45.0	792	94.0	673	3.3	747	39.2	792	90.9
683	11.1	748	57.1	795	97.5	683	5.9	748	46.1	795	95.6	683	3.9	748	40.7	795	92.7
686	12.3	749	60.9	803	98.0	686	7.0	749	49.7	803	96.6	686	4.5	749	43.7	803	95.5
695	13.7	751	62.0	807	98.9	695	7.3	751	51.7	807	98.4	695	4.8	751	45.3	807	97.9
702	14.5	752	63.9	820	100.0	702	7.9	752	53.8	820	100.0	702	5.8	752	47.8	820	100.0
703	16.5	754	64.8			703	8.9	754	54.9			703	6.1	754	49.4		
706	18.0	755	67.4			706	9.3	755	57.3			706	7.6	755	53.3		
712	20.2	757	71.4			712	11.0	757	61.0			712	8.7	757	55.8		
714	21.8	759	72.7			714	11.6	759	62.7			714	9.9	759	56.8		
719	24.8	760	74.1			719	14.3	760	64.3			719	11.3	760	57.8		
720	26.4	761	75.3			720	15.6	761	65.9			720	12.6	761	58.7		
724	28.5	762	76.6			724	17.1	762	67.4			724	13.1	762	59.6		
725	30.9	763	77.5			725	19.0	763	68.6			725	15.2	763	61.6		
728	32.2	764	78.8			728	20.0	764	70.5			728	16.5	764	62.0		
729	32.9	766	79.5			729	21.0	766	72.3			729	18.1	766	63.4		
730	34.2	767	80.9			730	22.4	767	74.3			730	19.1	767	65.8		
732	36.2	768	82.5			732	23.5	768	75.8			732	20.0	768	67.8		
733	37.6	770	83.5			733	24.4	770	76.9			733	20.9	770	69.7		
734	39.0	771	84.3			734	26.0	771	77.8			734	22.5	771	71.6		
735	41.0	772	86.0			735	27.5	772	79.3			735	23.2	772	73.3		
736	42.0	774	87.3			736	28.4	774	81.2			736	24.1	774	75.1		
738	44.5	776	88.2			738	31.6	776	83.0			738	26.8	776	77.5		
740	45.9	777	89.5			740	32.7	777	85.3			740	27.7	777	78.8		
741	48.6	779	90.3			741	35.8	779	86.7			741	29.7	779	79.8		
743	49.4	781	91.4			743	37.6	781	87.6			743	31.5	781	82.0		
744	51.9	784	93.8			744	39.6	784	90.7			744	32.7	784	85.3		
745	53.1	787	94.6			745	41.5	787	92.2			745	35.0	787	88.0		

Table 13

Writing: Cumulative Scale Score Distributions for Field Tests by Grade

Grade K				Grade 01				Grade 02				Grade 03								Grade 04					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
515	50.9	680	99.8	515	10.5	659	86.3	515	3.6	659	38.7	515	1.5	637	8.0	670	27.6	730	94.8	575	4.2	668	22.8	720	75.6
559	58.6	683	99.9	559	13.1	663	89.4	559	4.5	663	43.9	551	1.5	638	8.1	671	29.8	739	96.2	595	4.9	669	23.8	721	76.9
567	68.4	685	100.0	567	16.4	666	90.7	567	5.1	666	46.2	556	1.9	639	8.6	673	30.8	743	97.0	596	5.2	672	25.3	723	77.9
572	76.6			572	19.3	669	91.8	572	5.7	669	49.6	562	2.0	640	8.9	674	32.0	746	98.4	602	5.6	673	26.3	724	79.6
588	82.4			588	21.1	670	92.4	588	6.2	670	52.7	578	2.2	642	9.1	675	33.4	775	100.0	610	5.9	676	28.0	725	80.9
594	85.8			594	24.3	673	93.3	594	7.4	673	54.8	582	2.5	643	10.3	676	34.7			611	6.5	677	28.6	728	82.7
597	88.6			597	26.5	676	94.2	597	7.8	676	58.7	585	2.7	645	10.6	677	36.9			615	6.9	680	30.2	729	84.2
605	89.4			605	28.8	677	94.8	605	8.0	677	62.8	593	2.9	646	10.9	678	38.8			620	7.6	681	31.9	730	85.3
609	90.5			609	31.7	680	95.5	609	9.0	680	65.8	596	3.2	647	11.3	680	40.2			621	7.7	683	32.7	733	86.7
611	92.1			611	34.5	683	96.1	611	9.2	683	68.3	598	3.5	648	11.6	681	41.6			624	8.1	684	34.1	735	88.9
616	93.0			616	37.0	685	96.9	616	9.8	685	72.5	603	3.6	649	12.0	682	43.8			628	8.3	685	35.4	739	89.8
620	93.8			620	39.8	690	97.5	620	10.6	690	75.1	606	3.7	650	12.2	684	47.0			629	9.0	687	38.3	741	91.9
622	94.7			622	43.1	692	97.8	622	11.1	692	77.9	607	3.9	651	12.9	686	49.5			631	9.3	689	39.4	745	92.7
625	95.2			625	45.4	695	98.4	625	11.8	695	81.4	611	4.2	652	13.7	688	53.0			635	9.9	691	42.1	747	95.0
628	95.8			628	48.4	703	98.7	628	12.5	703	83.9	613	4.4	653	14.4	690	55.4			636	10.5	693	43.4	753	95.7
630	96.7			630	52.4	705	99.0	630	13.4	705	87.2	614	4.6	654	14.8	692	60.1			637	11.1	695	46.0	755	97.0
633	97.3			633	54.6	708	99.3	633	14.3	708	90.0	617	4.9	655	15.3	694	62.8			641	11.5	696	47.4	762	97.9
636	97.5			636	57.3	730	100.0	636	16.3	730	100.0	620	5.1	656	16.1	697	67.3			642	12.1	699	50.7	764	98.1
638	97.8			638	61.2			638	17.5			622	5.3	657	16.8	699	70.0			646	12.7	700	52.2	765	98.5
640	98.0			640	64.3			640	18.9			624	5.4	658	17.8	702	75.3			647	13.3	703	55.9	774	98.7
643	98.3			643	66.6			643	21.2			625	5.9	659	18.7	704	77.3			651	13.7	704	57.3	776	99.2
644	98.9			644	70.2			644	23.1			627	6.0	660	19.4	708	80.4			652	15.0	707	60.7	778	99.3
646	99.0			646	72.3			646	24.7			629	6.3	662	22.0	711	82.8			656	16.2	708	61.9	795	99.5
650	99.3			650	74.6			650	27.7			631	6.7	663	22.5	715	84.9			657	16.7	711	66.0	798	99.7
657	99.4			651	78.6			651	29.4			632	6.8	664	23.3	716	86.8			660	17.9	712	67.4	800	99.8
663	99.5			653	79.8			653	31.1			633	7.1	665	24.1	719	89.4			661	18.4	715	70.5	825	100.0
669	99.6			656	82.3			656	34.0			635	7.5	667	24.9	724	91.1			664	19.7	717	72.6		
670	99.8			657	85.2			657	37.0			636	7.8	668	26.4	727	93.0			665	20.5	719	73.8		

Table 13 (cont.)

Writing: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 05						Grade 06						Grade 07						Grade 08						Grade 09					
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
575	5.3	669	23.7	723	69.5	575	3.6	669	21.7	723	68.9	580	4.3	692	26.9	754	83.0	580	2.9	692	30.5	754	77.9	580	2.9	692	30.5	754	77.9
595	5.7	672	25.3	724	71.5	595	4.1	672	23.6	724	71.1	587	4.9	693	27.8	756	84.0	587	3.6	693	31.9	756	78.7	587	3.6	693	31.9	756	78.7
596	5.9	673	25.7	725	73.5	596	4.8	673	24.1	725	72.8	590	5.1	694	28.2	757	85.3	590	3.9	694	32.9	757	80.9	590	3.9	694	32.9	757	80.9
602	6.3	676	26.7	728	74.8	602	5.2	676	25.4	728	74.6	606	6.2	697	30.5	761	86.3	606	4.8	697	34.7	761	82.0	606	4.8	697	34.7	761	82.0
610	7.0	677	27.3	729	76.7	610	5.3	677	26.0	729	76.4	607	6.7	698	31.2	762	87.1	607	5.0	698	35.4	762	83.5	607	5.0	698	35.4	762	83.5
611	7.4	680	28.6	730	78.1	611	5.7	680	28.2	730	78.2	618	7.2	701	32.3	764	88.9	618	5.5	701	36.0	764	85.2	618	5.5	701	36.0	764	85.2
615	8.0	681	29.3	733	79.5	615	5.9	681	29.9	733	79.7	619	7.7	702	34.0	768	90.6	619	6.3	702	38.4	768	86.8	619	6.3	702	38.4	768	86.8
620	8.9	683	29.8	735	83.6	620	6.3	683	30.6	735	82.3	620	8.1	706	36.6	769	91.3	620	7.1	706	40.3	769	87.5	620	7.1	706	40.3	769	87.5
621	9.1	684	30.8	739	84.5	621	6.9	684	31.7	739	83.9	628	8.3	707	37.4	772	92.8	628	7.9	707	41.8	772	89.2	628	7.9	707	41.8	772	89.2
624	9.4	685	31.4	741	87.0	624	7.0	685	32.8	741	86.1	629	9.1	710	39.5	778	95.1	629	8.8	710	44.0	778	91.9	629	8.8	710	44.0	778	91.9
628	10.1	687	33.0	745	88.1	628	7.4	687	34.1	745	87.7	636	10.0	711	40.7	783	96.0	636	9.0	711	45.2	783	93.8	636	9.0	711	45.2	783	93.8
629	10.6	689	34.1	747	90.9	629	7.5	689	35.9	747	89.5	638	10.9	714	43.2	789	96.4	638	10.3	714	47.8	789	94.5	638	10.3	714	47.8	789	94.5
631	10.7	691	37.0	753	91.9	631	7.7	691	37.4	753	90.6	643	11.7	715	44.1	792	97.3	643	11.0	715	48.7	792	95.7	643	11.0	715	48.7	792	95.7
635	11.1	693	38.2	755	94.5	635	8.4	693	38.8	755	92.5	646	12.8	718	46.0	798	97.7	646	12.6	718	49.5	798	96.4	646	12.6	718	49.5	798	96.4
636	11.4	695	40.6	762	95.5	636	9.1	695	40.5	762	93.2	650	12.9	719	49.1	806	97.9	650	13.2	719	51.8	806	97.1	650	13.2	719	51.8	806	97.1
637	11.8	696	41.5	764	96.4	637	9.6	696	41.8	764	94.2	653	14.1	723	53.6	814	98.7	653	14.3	723	55.4	814	97.7	653	14.3	723	55.4	814	97.7
641	12.3	699	44.3	765	97.4	641	10.2	699	44.0	765	95.3	656	14.6	727	55.8	821	99.2	656	14.6	727	58.5	821	98.7	656	14.6	727	58.5	821	98.7
642	13.2	700	45.8	774	97.6	642	10.9	700	46.1	774	96.1	660	15.6	728	58.0	845	100.0	660	16.4	728	59.7	845	100.0	660	16.4	728	59.7	845	100.0
646	13.9	703	47.7	776	98.3	646	11.7	703	48.2	776	97.0	662	16.0	731	61.2			662	16.8	731	62.1			662	16.8	731	62.1		
647	15.0	704	49.6	778	98.6	647	13.1	704	49.5	778	97.8	666	16.5	732	64.0			666	17.7	732	63.3			666	17.7	732	63.3		
651	15.5	707	52.4	795	98.8	651	14.1	707	52.2	795	98.3	667	17.0	736	67.3			667	19.0	736	65.3			667	19.0	736	65.3		
652	17.0	708	53.7	798	99.2	652	15.2	708	53.3	798	98.6	668	17.7	737	69.8			668	19.5	737	66.8			668	19.5	737	66.8		
656	18.0	711	56.1	800	99.4	656	15.9	711	56.8	800	99.2	672	17.9	740	72.7			672	20.1	740	70.0			672	20.1	740	70.0		
657	18.4	712	57.8	825	100.0	657	16.6	712	57.7	825	100.0	673	19.3	742	74.3			673	22.0	742	70.9			673	22.0	742	70.9		
660	19.9	715	61.2			660	17.2	715	60.1			678	21.3	745	75.3			678	24.5	745	71.9			678	24.5	745	71.9		
661	20.6	717	62.7			661	18.2	717	62.6			683	22.7	746	76.5			683	26.1	746	73.4			683	26.1	746	73.4		
664	21.7	719	64.6			664	19.1	719	64.0			684	23.7	748	78.2			684	26.9	748	74.1			684	26.9	748	74.1		
665	22.1	720	66.4			665	20.0	720	66.0			688	25.3	750	79.8			688	28.4	750	75.6			688	28.4	750	75.6		
668	23.3	721	68.6			668	21.4	721	67.4			689	26.1	751	81.6			689	29.8	751	76.8			689	29.8	751	76.8		

Table 13 (cont.)

Writing: Cumulative Scale Score Distributions for Field Tests by Grade

Grade 10							
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
600	3.3	678	21.6	714	54.3	776	95.4
607	3.5	679	22.5	715	55.1	781	96.3
608	3.7	680	23.1	717	57.1	786	96.6
618	4.1	682	23.8	719	58.6	790	97.2
621	4.5	683	25.6	720	59.2	798	97.9
629	4.6	684	26.8	721	62.1	806	98.4
630	5.5	686	28.7	723	63.7	814	99.0
637	6.0	688	29.8	724	64.7	828	99.4
638	6.4	689	31.1	726	66.8	850	100.0
643	6.6	690	32.2	728	69.0		
644	7.7	691	33.1	730	70.6		
649	7.9	693	34.0	732	71.9		
650	9.2	694	35.6	733	73.6		
654	9.4	695	36.6	735	76.0		
655	9.9	697	37.6	738	78.5		
656	10.7	698	38.9	741	80.8		
659	11.3	699	39.4	743	83.1		
660	12.4	700	40.3	747	84.4		
661	13.1	702	41.6	749	85.5		
663	13.9	703	42.4	750	86.3		
665	14.6	704	42.9	753	88.0		
666	15.4	706	44.9	756	89.0		
668	15.7	707	45.8	757	90.2		
669	16.7	708	46.7	761	91.6		
670	17.9	709	49.2	764	92.5		
672	18.0	710	50.4	766	93.4		
674	19.7	712	51.6	770	94.2		
676	20.0	713	53.1	773	94.4		

Grade 11							
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
600	3.4	683	16.3	720	53.1	798	96.7
607	3.5	684	17.4	721	54.3	806	97.6
608	3.6	686	19.1	723	55.4	814	98.6
618	3.9	688	19.8	724	56.8	828	98.9
629	4.1	689	20.8	726	58.8	850	100.0
630	4.5	690	22.3	728	61.4		
643	4.8	691	22.7	730	63.0		
644	5.2	693	23.5	732	65.0		
649	5.4	694	24.8	733	66.9		
650	6.2	695	25.8	735	68.6		
655	6.3	697	26.4	738	71.7		
656	6.9	698	27.7	741	73.5		
659	7.1	699	28.8	743	76.7		
660	7.4	700	29.4	747	78.6		
661	8.1	702	31.0	749	80.2		
663	8.3	703	32.8	750	82.0		
665	8.6	704	34.0	753	83.9		
666	9.0	706	35.6	756	85.5		
668	9.9	707	37.3	757	87.1		
669	10.3	708	38.7	761	87.6		
670	10.8	709	40.6	764	88.3		
672	11.2	710	41.8	766	90.2		
674	12.4	712	42.7	770	91.1		
676	12.8	713	45.1	773	92.2		
678	13.1	714	46.8	776	93.9		
679	14.1	715	48.6	781	94.8		
680	14.8	717	50.4	786	95.6		
682	15.4	719	51.8	790	96.0		

Grade 12							
SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt	SS	cum. pcnt
600	2.5	684	13.8	721	50.1	806	96.8
607	2.6	686	14.6	723	52.7	814	97.8
621	2.7	688	15.5	724	53.7	828	98.6
630	2.8	689	16.3	726	54.6	850	100.0
637	2.9	690	17.1	728	58.5		
638	3.2	691	17.7	730	59.7		
644	3.3	693	18.5	732	61.5		
650	3.8	694	19.5	733	62.9		
654	3.9	695	20.3	735	64.5		
655	4.0	697	21.4	738	68.2		
656	4.2	698	22.4	741	69.3		
659	4.5	699	23.8	743	72.1		
660	4.8	700	25.3	747	73.6		
661	5.0	702	26.3	749	74.7		
663	5.3	703	27.9	750	76.8		
665	5.9	704	29.2	753	78.8		
666	6.5	706	30.9	756	80.5		
668	7.1	707	32.7	757	82.1		
669	7.2	708	34.2	761	83.3		
670	7.7	709	35.3	764	84.8		
672	8.2	710	37.1	766	86.8		
674	9.5	712	38.5	770	87.7		
676	9.9	713	40.0	773	89.2		
678	10.8	714	42.0	776	90.9		
679	10.9	715	43.8	781	92.0		
680	12.1	717	44.6	786	93.7		
682	12.8	719	47.2	790	94.8		
683	13.3	720	48.5	798	95.5		

Table 14

Listening/Speaking: Percentage of Field Test Students Reaching Benchmarks

	Grade K	Grade 01	Grade 02	Grade 03	Grade 04	Grade 05	Grade 06
<620	42.8	12.5	6.3	5.0	3.9	6.2	4.8
620<=SS<660	44.4	44.9	21.5	11.8	9.0	9.0	10.1
660<=SS<700	11.6	36.5	51.7	44.5	31.7	26.6	29.1
700<=SS<740	1.1	6.0	18.9	33.7	47.0	45.9	39.9
740<=SS	0.0	0.2	1.6	5.0	8.3	12.2	16.2

	Grade 07	Grade 08	Grade 09	Grade 10	Grade 11	Grade 12
<620	8.4	7.2	10.8	3.9	1.7	0.6
620<=SS<660	7.8	8.6	12.4	8.6	4.3	3.3
660<=SS<700	24.3	20.9	26.5	23.3	22.1	21.5
700<=SS<740	54.5	54.9	43.7	49.4	54.4	52.1
740<=SS	5.0	8.4	6.7	14.8	17.6	22.5

Table 15

Reading: Percentage of Field Test Students Reaching Benchmarks

	Grade K	Grade 01	Grade 02	Grade 03	Grade 04	Grade 05	Grade 06
<620	98.8	78.9	25.2	11.1	11.2	10.8	8.7
620<=SS<660	0.5	12.0	25.1	14.0	10.9	9.8	10.7
660<=SS<700	0.6	5.7	27.5	38.1	29.4	22.2	20.3
700<=SS<740	0.1	2.7	15.0	29.9	39.6	41.6	40.5
740<=SS<780	0.0	0.5	5.5	5.2	7.9	14.5	16.8
780<=SS	0.0	0.3	1.7	1.7	1.0	1.2	3.0

	Grade 07	Grade 08	Grade 09	Grade 10	Grade 11	Grade 12
<620	13.7	9.6	11.2	9.1	4.8	3.1
620<=SS<660	3.1	2.5	2.5	0.0	0.0	0.0
660<=SS<700	10.7	8.6	10.7	4.6	2.5	1.8
700<=SS<740	41.6	38.6	35.5	30.8	24.3	22.0
740<=SS<780	27.4	34.6	33.9	45.8	55.1	53.0
780<=SS	3.5	6.1	6.2	9.7	13.3	20.2

Table 16

Writing: Percentage of Field Test Students Reaching Benchmarks

	Grade K	Grade 01	Grade 02	Grade 03	Grade 04	Grade 05	Grade 06
<620	93.0	37.0	9.8	4.9	6.9	8.0	5.9
620<=SS<660	6.5	49.3	28.9	13.8	9.7	10.4	10.8
660<=SS<700	0.6	12.1	42.7	51.3	34.0	25.8	27.4
700<=SS<740	0.0	1.6	18.6	26.2	39.1	40.3	39.9
740<=SS<780	0.0	0.0	0.0	3.8	9.5	14.0	13.9
780<=SS	0.0	0.0	0.0	0.0	0.7	1.4	2.2

	Grade 07	Grade 08	Grade 09	Grade 10	Grade 11	Grade 12
<620	7.7	5.6	6.3	4.1	3.9	2.6
620<=SS<660	6.9	5.3	8.2	7.2	3.2	1.8
660<=SS<700	16.6	17.1	20.8	28.0	21.7	19.3
700<=SS<740	38.6	34.1	31.4	39.2	42.8	44.3
740<=SS<780	25.3	29.7	25.2	16.8	22.3	22.7
780<=SS	4.9	8.3	8.1	4.6	6.1	9.1

Table 17

Item Statistics by Modality, Test Level, and Final Form

Modality	Test Level/ Form	No. of Items	Difficulty		Discrimination	
			Mean	SD	Mean	SD
Listening	A1	20	0.74	0.13	0.62	0.09
	A2	20	0.74	0.13	0.57	0.09
	B1	22	0.72	0.13	0.54	0.09
	B2	22	0.69	0.15	0.54	0.07
	C1	22	0.73	0.12	0.61	0.09
	C2	22	0.71	0.13	0.60	0.08
	D1	22	0.76	0.14	0.58	0.09
	D2	22	0.75	0.15	0.55	0.09
Speaking	A1	10	0.72	0.14	0.75	0.08
	A2	10	0.68	0.14	0.74	0.08
	B1	13	0.78	0.08	0.75	0.10
	B2	13	0.80	0.08	0.75	0.10
	C1	13	0.69	0.11	0.74	0.13
	C2	13	0.67	0.11	0.75	0.12
	D1	13	0.72	0.11	0.72	0.13
	D2	13	0.74	0.07	0.71	0.13
Reading	A1	21	0.70	0.16	0.64	0.13
	A2	21	0.70	0.16	0.63	0.16
	A1+Ext	31	0.64	0.17	0.58	0.14
	A2+Ext	31	0.66	0.16	0.61	0.14
	B1	27	0.60	0.16	0.59	0.14
	B2	27	0.61	0.13	0.60	0.10
	C1	26	0.57	0.14	0.59	0.11
	C2	26	0.57	0.13	0.58	0.09
	D1	26	0.62	0.13	0.57	0.10
	D2	26	0.63	0.12	0.57	0.08
Writing	A1	7	0.50	0.13	0.84	0.05
	A2	7	0.47	0.12	0.80	0.01
	A1+Ext	16	0.58	0.14	0.77	0.12
	A2+Ext	16	0.57	0.14	0.76	0.10
	B1	25	0.60	0.11	0.58	0.14
	B2	25	0.61	0.11	0.59	0.15
	C1	25	0.62	0.13	0.57	0.19
	C2	25	0.59	0.14	0.58	0.17
	D1	25	0.63	0.14	0.65	0.12
	D2	25	0.67	0.11	0.61	0.12

Table 18

Scoring Table for Locator Test

Raw Score	Scale Score	SEM
0	345	200
1	345	200
2	345	200
3	345	200
4	345	200
5	345	200
6	652	57
7	676	38
8	692	24
9	702	17
10	710	13
11	717	12
12	723	11
13	729	10
14	735	10
15	742	11
16	750	12
17	762	16
18	820	74

Note: SEMs for scale scores of 345 and 820 are approximate.

Table 19

Proportions of Students in the Field Test Scoring Below the Locator Cut Points and At or Above the Locator Cut Points

Test Level	Cut Points		
	Below 652	At or Above 652	Total
A	0.59	0.41	1.00
B	0.17	0.83	1.00

Test Level	Cut Points		
	Below 702	At or Above 702	Total
B	0.48	0.52	1.00
C	0.26	0.74	1.00

Test Level	Cut Points		
	Below 729	At or Above 729	Total
C	0.50	0.50	1.00
D	0.23	0.77	1.00

Figure 1. Listening/Speaking: Test characteristic curves for field test forms

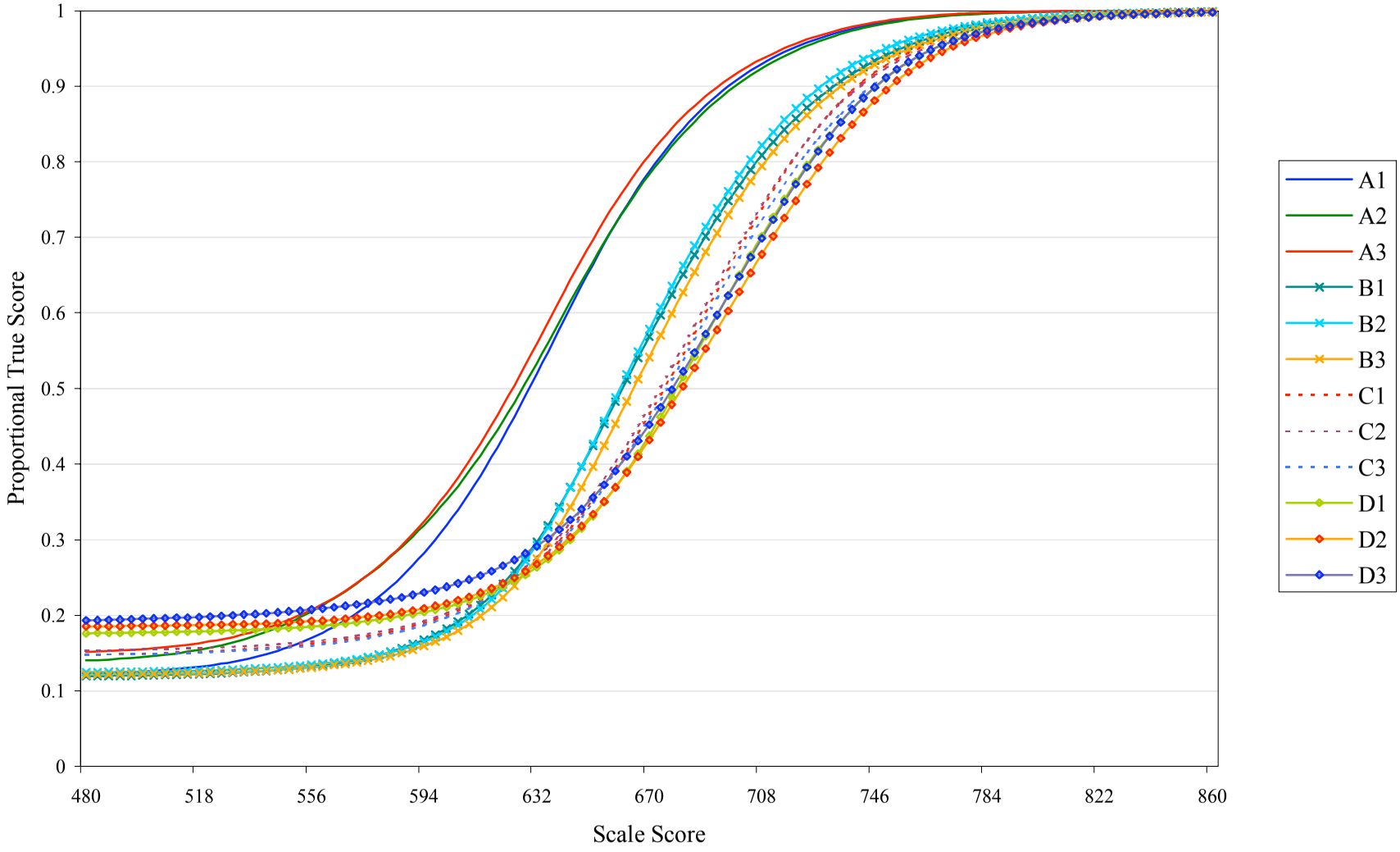


Figure 2. Reading: Test characteristic curves for field test forms

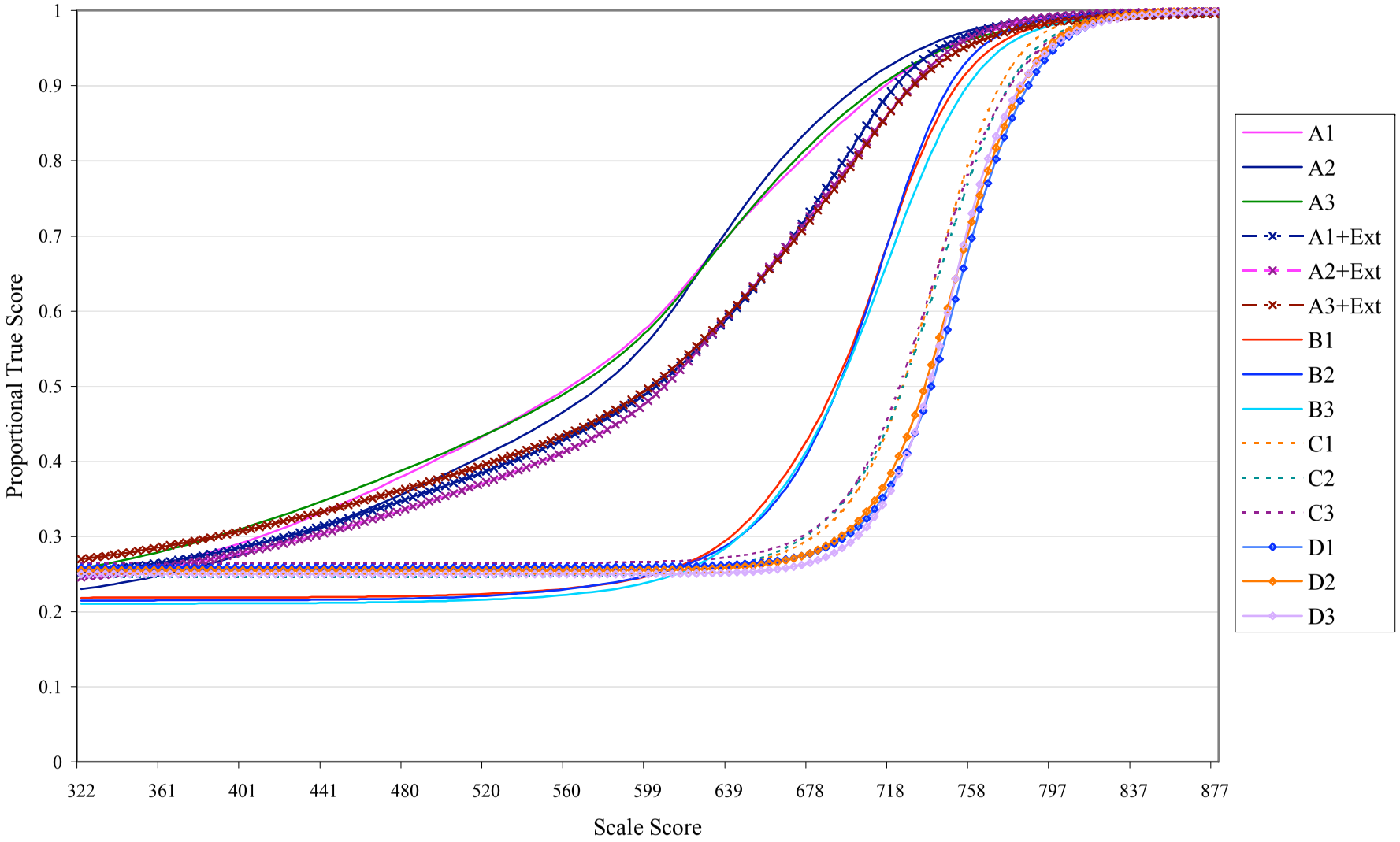


Figure 3. Writing: Test characteristic curves for field test forms

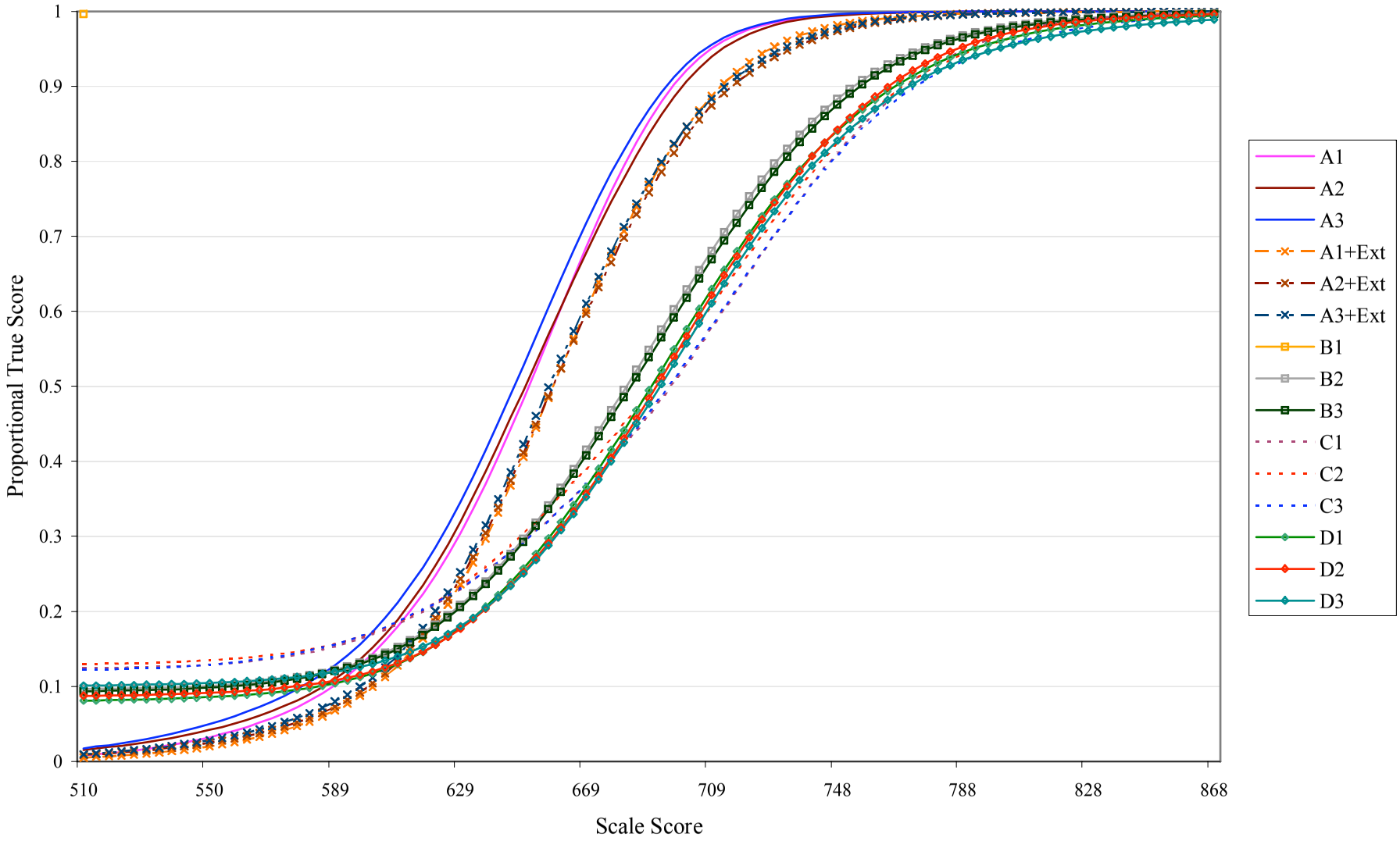


Figure 4. Listening/Speaking: Standard errors of measurement for field test forms

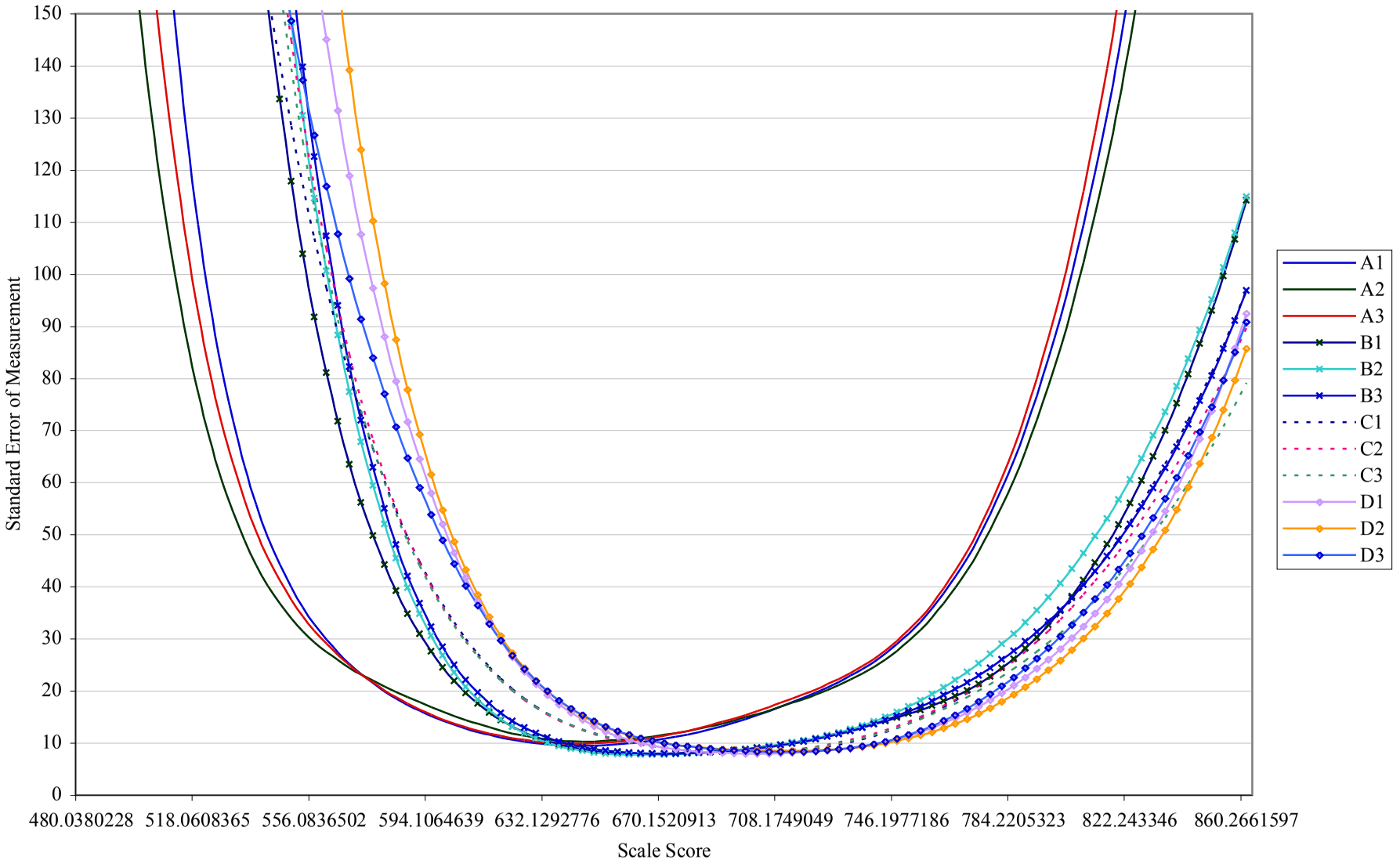


Figure 5. Reading: Standard errors of measurement for field test forms

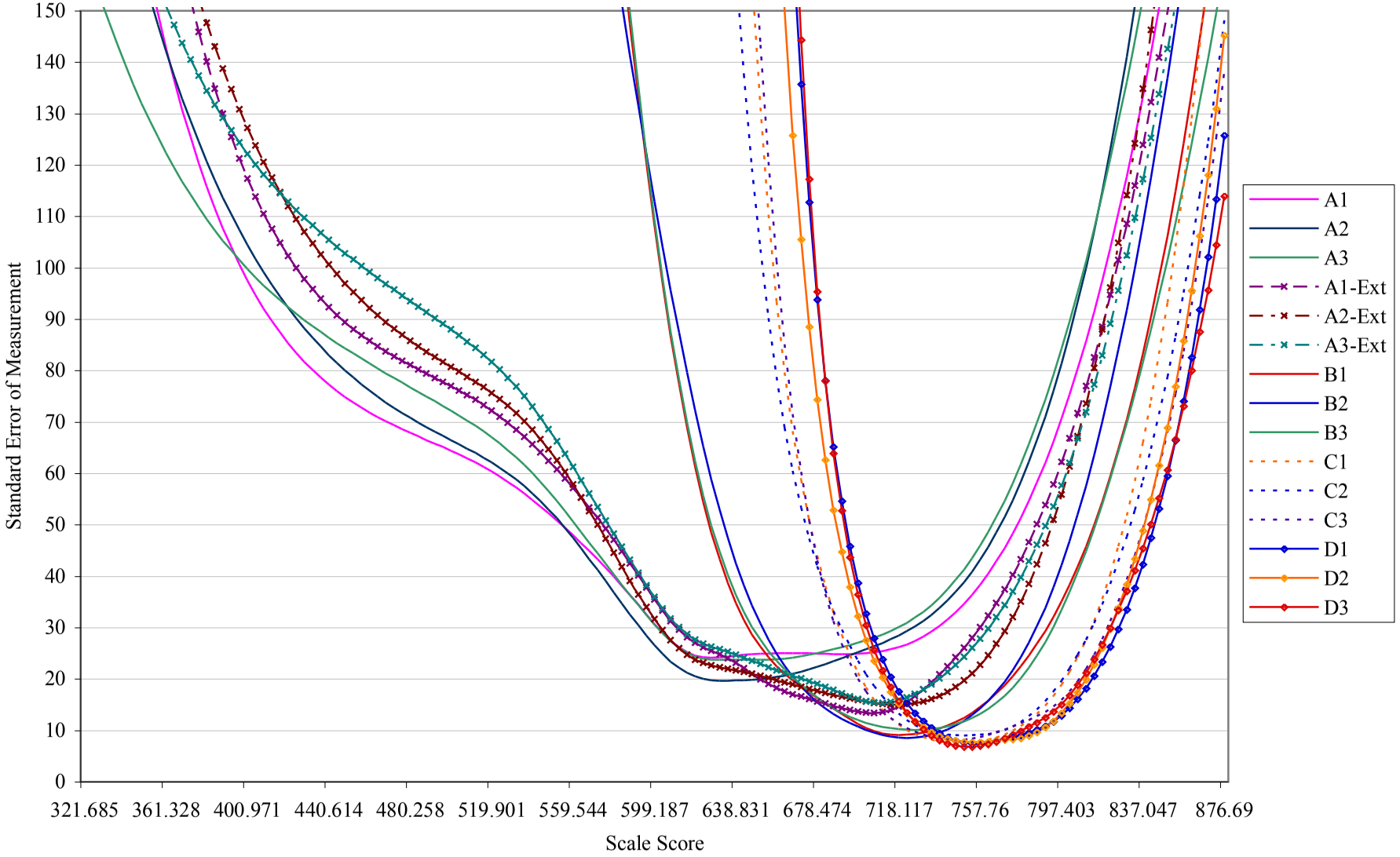


Figure 6. Writing: Standard errors of measurement for field test forms

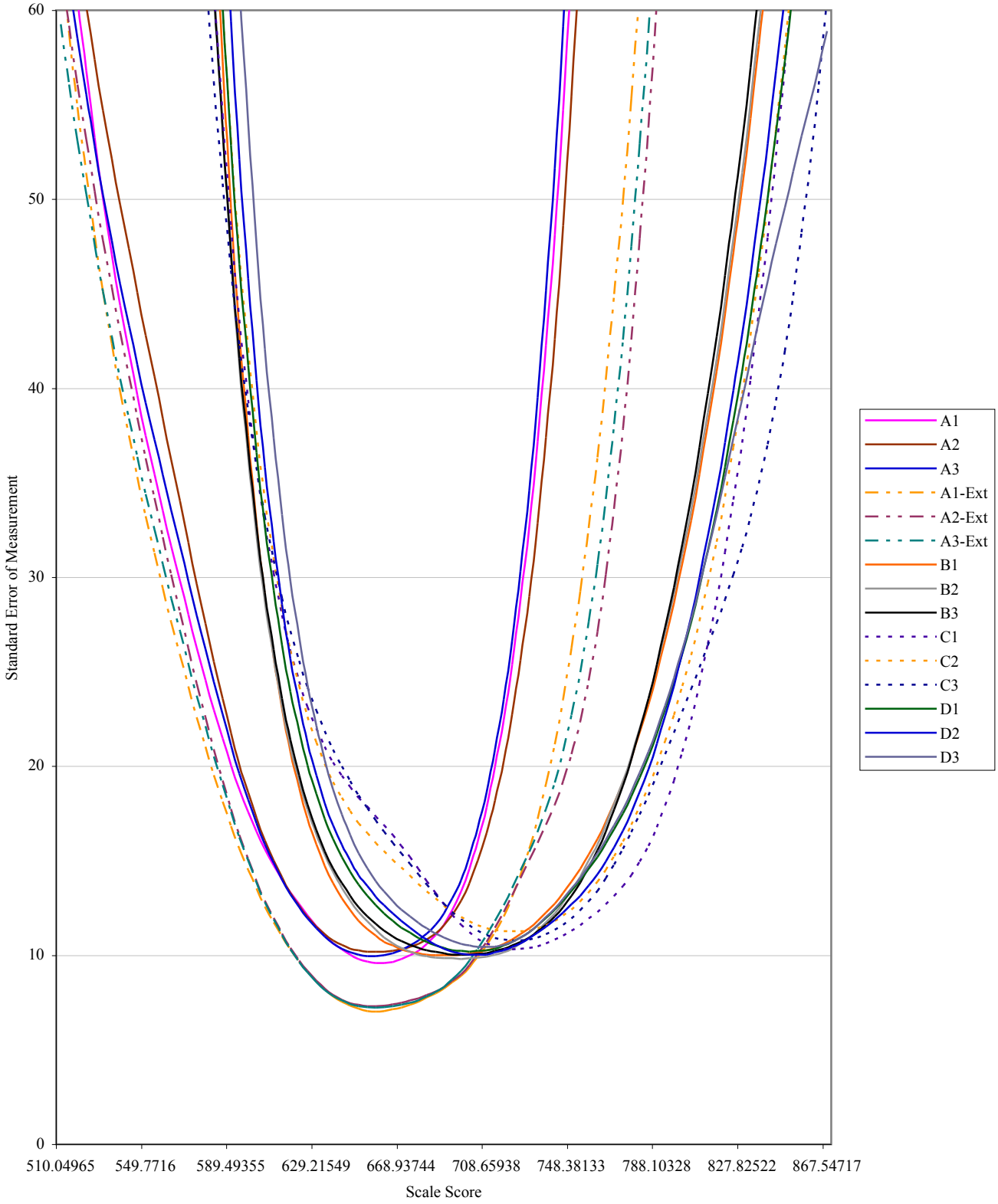


Figure 7. Listening/Speaking: Cumulative percentage scale score distributions by grade based on field test forms

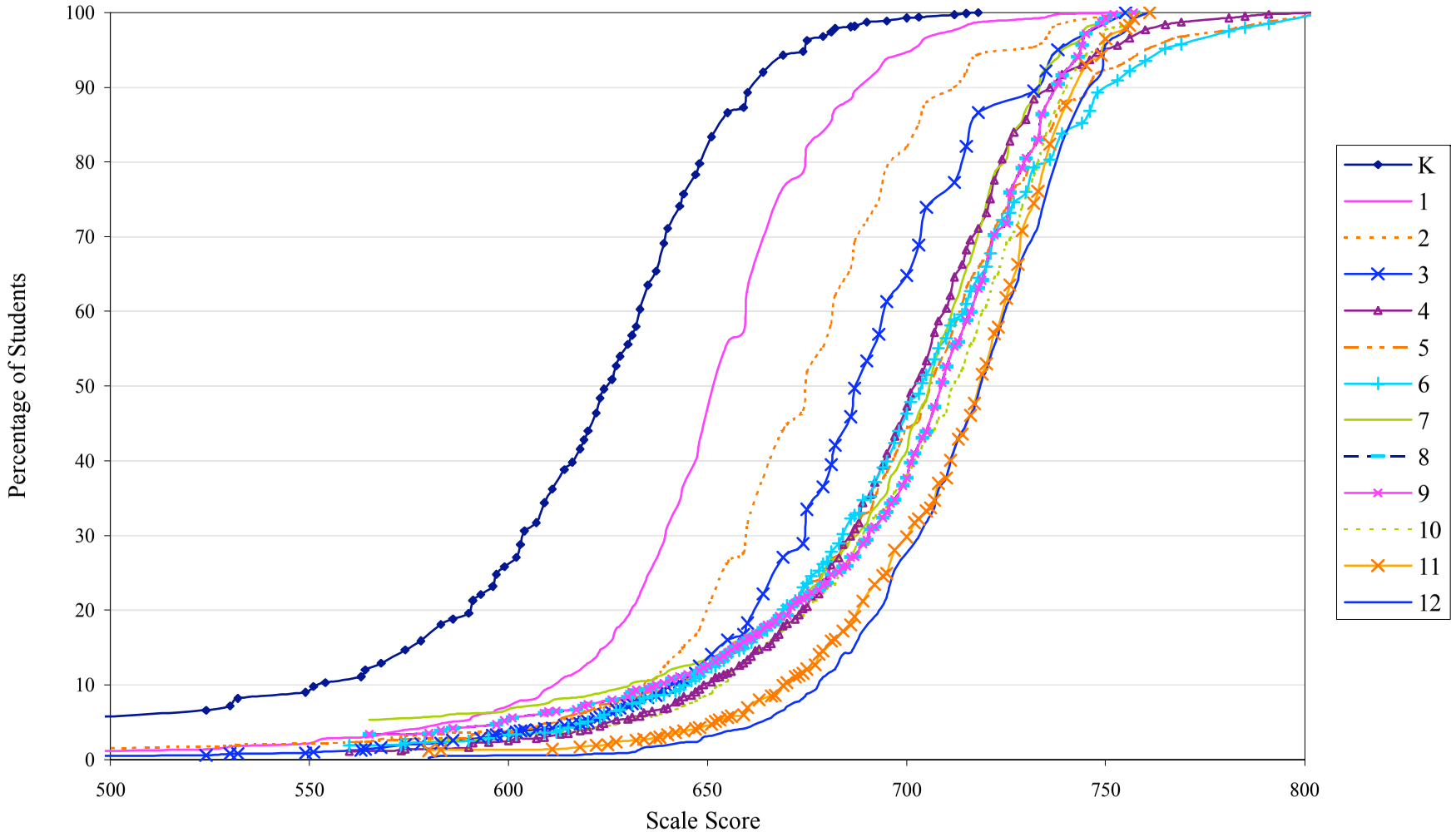


Figure 8. Reading: Cumulative percentage scale score distributions by grade based on field test forms

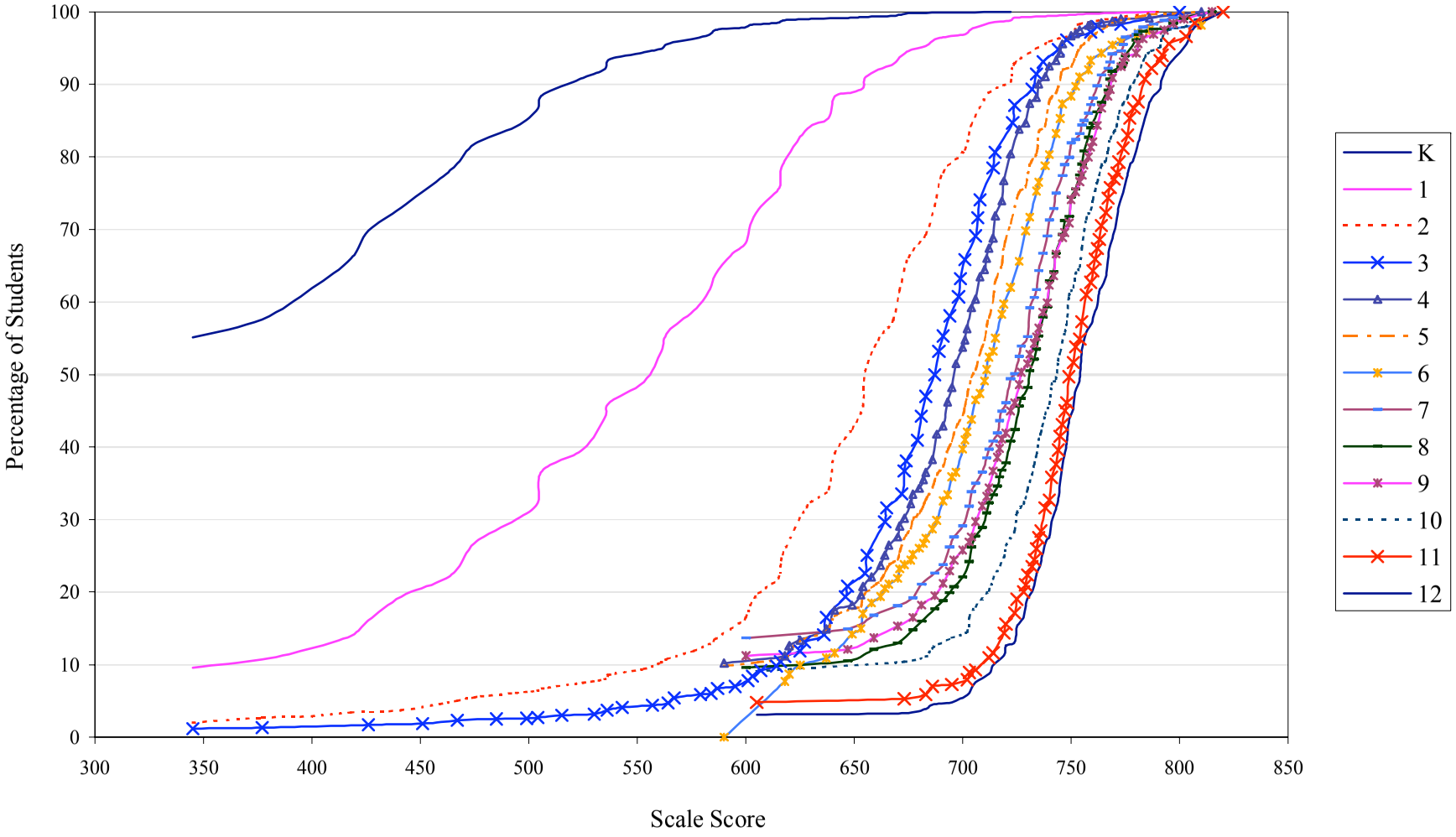
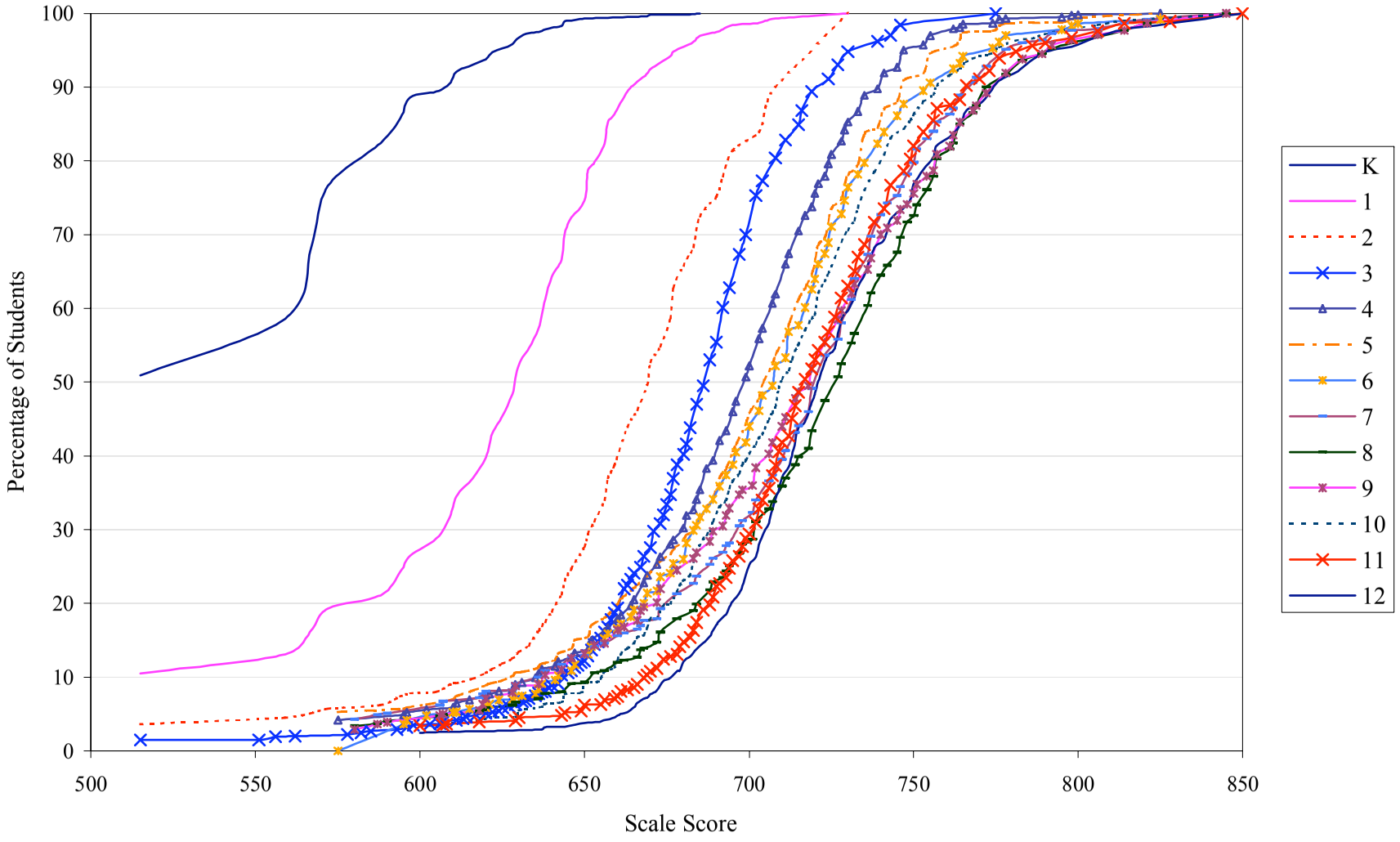


Figure 9. Writing: Cumulative percentage scale score distributions by grade based on field test forms



Appendix B: Overview of Skills Tested

CELLA Final Test Forms Overview—Level A and Level A Extension (Ax)

Modality	Benchmarks	Level		Item Types
		A	Ax	
Listening	Vocabulary	20-30%	N/A	Listening Vocabulary (CR)
	Comprehension	70-80%	N/A	Listen and Match
				Short Talks (Teacher Instruction)
				Extended Listening Comprehension
Speaking	Pronunciation* Vocabulary Grammar/Sentence* Discourse	0-15% 20-30% 0-20% 45-65%	N/A N/A N/A N/A	Oral Vocabulary (CR)
				Speech Functions (CR)
				Personal Opinion (CR)
				Story Retelling (CR)
Reading	Print Concepts	15-30%	10-25%	Demonstrating Print Concepts (CR)
				Naming Letters (CR)
	Phonemic Awareness, Decoding, and Word Recognition	40-55%	25-40%	Reading Aloud for Fluency (CR)
				Listen, Read, and Match
	Vocabulary	4-15%	5-15%	Listen, Read, and Match (antonym)
				Vocabulary in Context
	Comprehension	15-30%	35-45%	Short Reading Comprehension
				Reading Comprehension Set
Writing	Writing from Dictation	10-25%	5-15%	Dictated Letters (CR)
	Spelling	30-50%	15-25%	Dictated Words (CR)
	Punctuation & Capitalization**	0-15%	5-20%	Dictated Sentences (CR)
	Grammar**	0-20%	15%	Descriptive Sentences (CR)
	Writing Sentences	20-40%	15-30%	
	Writing Paragraphs	--	10-20%	Multiple Sentences (CR)
	Editing	--	3-15%	Editing

(CR) = Constructed Response Items
*Pronunciation and Grammar Speaking Benchmarks are assessed through the scoring rubric for Discourse items. However, all Discourse test items are mapped to Discourse Benchmarks (since all items are mapped to a single Benchmark). Thus, though the alignment mappings may appear to suggest otherwise, the Pronunciation and Grammar Benchmarks in Speaking are assessed.
**Grammar and Punctuation & Capitalization Writing Benchmarks are assessed through some of the scoring rubrics for Sentences in Level A and the scoring rubrics for Sentences and Paragraphs in Level A Extension and Levels B - D. Thus, though the alignment mappings may appear to suggest otherwise in some cases, Grammar and Punctuation & Capitalization Benchmarks are assessed at all levels. (See CELLA Scoring Guide for Writing, Level A.)

CELLA Final Test Forms Overview—Levels B, C, and D

Modality	Benchmarks	LEVEL			Item Type
		B	C	D	
Listening	Vocabulary	30-40%	30-40%	30-40%	Listen and Match
					Picture Description
	Comprehension	60-70%	60-70%	60-70%	Short Talks (Teacher Instruction)
					Short Talks (Partial Dialogue)
					Extended Listening Comprehension
Speaking	Vocabulary Pronunciation* Grammar/Sentence Discourse	15-30% 0-15% 0-20% 55-75%	15-30% 0-15% 0-20% 55-75%	15-30% 0-15% 0-20% 55-75%	Oral Vocabulary (CR)
					Speech Functions (CR)
					Personal Opinion (CR)
					Story Retelling (CR)
					Graph Interpretation (CR)
Reading	Vocabulary	30-40%	30-40%	30-40%	Discrete Vocabulary
					Vocabulary in Context
	Comprehension	60-70%	60-70%	60-70%	Reading Aloud for Fluency (CR)
					Reading Comprehension
Writing	Grammar**	15-30%	15-30%	15-30%	Grammar, Structure, Word Choice
	Punctuation & Capitalization**	3-15%	3-15%	3-15%	Grammar, Structure, Word Choice
					Recognizing Errors
	Editing, including Spelling	10-20%	10-20%	10-20%	Recognizing Errors
	Writing Sentences	25-40%	25-40%	30-40%	Writing Sentences (CR)
					Grammar, Structure, Word Choice
	Writing Paragraphs	25-40%	25-40%	30-40%	Writing Paragraphs (CR)
					Paragraph Choices

(CR) = Constructed Response Items
*Pronunciation and Grammar Speaking Benchmarks are assessed through the scoring rubric for Discourse items. However, all Discourse test items are mapped to Discourse Benchmarks (since all items are mapped to a single Benchmark). Thus, though the alignment mappings may appear to suggest otherwise, the Pronunciation and Grammar Benchmarks in Speaking are assessed.
** Grammar and Punctuation & Capitalization Writing Benchmarks are assessed through some of the scoring rubrics for Sentences and Paragraphs in Levels B - C. Thus, though the alignment mappings may appear to suggest otherwise in some cases, Grammar and Punctuation & Capitalization Benchmarks are assessed at all levels. (See CELLA Scoring Guide for Writing, Levels B and C.)

Appendix C: Scale Anchoring Proposal

Scale Anchoring Proposal

The CELLA project is designed to produce an assessment of English proficiency with four test levels, two forms per level in four skill areas (i.e., Reading, Writing, Listening and Speaking). Each skill area is assessed using a variety of multiple-choice and/or constructed-response items and three scales of measurement (i.e., Reading, Writing, and combined Listening/Speaking) will be created using item response theory. Moreover, the three scales of measurement will be linked vertically such that all four CELLA levels share a common scale.

Vertical scales can be used in longitudinal analysis of student and aggregate performance and allow for the possibility of functional level testing. However, scales of measurement are not readily interpretable in terms of what students can and cannot do. In order to provide such information, scale-anchoring procedures (Zwick, Senturk, Wang and Loomis, 2001; Beaton and Allen, 1992) take particular points on the vertical scale and attempt to provide this interpretive meaning. In particular, exemplar items that best reflect performance of groups of students that perform at different points on the vertical scale are identified using psychometric analyses. Content experts then synthesize these exemplar items into behavioral descriptions. This document describes how scale anchoring will be used to develop behavioral descriptions for the CELLA product.

Response Probability: In order to describe item/items reflecting what students at a particular point on a vertical scale can do, it is important to define first what the phrase "what students can do" means. To address this question, it is common to invoke the concept of "response probability" (RP). RP is defined as the probability of a group of students getting an item correct. It is also important to define the specific group of students used to calculate RP. In this study, we will identify 4 benchmarks for each vertical scale. Given that there are limited numbers of students participating in the field test, it is impossible to have enough students to score just at specific benchmarks. We introduce the 'benchmark interval' concept, which is a group of students' scores around specific benchmarks. The RP criterion is the percent correct that is required in order for the item/items to serve as exemplar items for students in a given score group. RP-65 and RP-74 have both been used by ETS for scale anchoring. Zwick, Senturk, Wang and Loomis (2001) showed that an RP greater than 70 is favored by a panel of science experts surveyed in the study. Although the Zwick (2001) study focused on item mapping for the 1996 National Assessment of Educational Progress (NAEP) science items, the panelists in the study were asked about the meaning of "what students can do" in a generic sense, and as a result the study provides some useful guidance for

scale anchoring. In the present study, we will use RP-74 for CELLA scale anchoring. This value is recommended by the Zwick (2001) study.

It is common practice to define different RP values for multiple-choice and constructed-response items because it is impossible for the respondents to guess the correct answer on the constructed-response items. To compensate for this, the most recent NAEP/ETS scale anchoring efforts have used a lower RP criterion for constructed-response items (RP_{CR}) than for the multiple-choice items (RP_{MC}). The relation between these two types of RP values, derived from the 3PL IRT model, was assumed to be as follows:

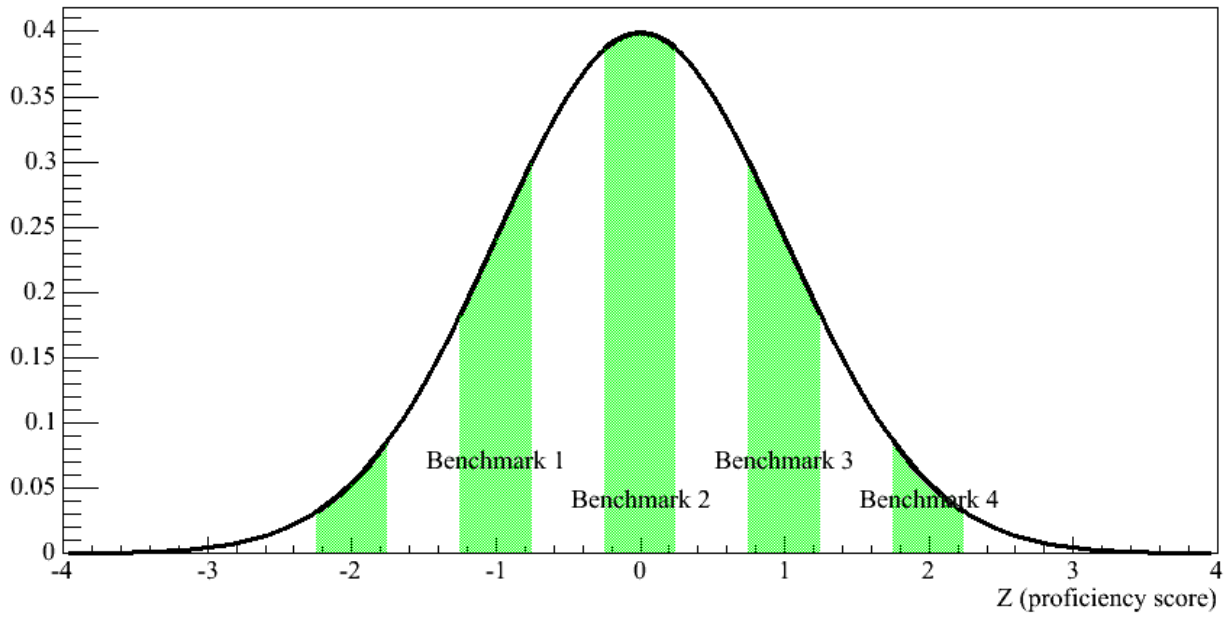
$RP_{MC} = c + (1-c)\ RP_{CR}$

Where the guessing parameter, c, was assumed to be .25. Following this precedent, we determined for the constructed-response items, a RP_{CR} value of .65, which will be used in this study. These RPs values are applied throughout the scale anchoring analyses described below.

Discrimination Criterion: In order to ensure that the exemplar items selected for a given benchmark exemplify only the given benchmark, not the benchmark below, a discrimination criterion is imposed. The discrimination criterion at a particular benchmark interval is defined as the difference between the item's value of *p* for that benchmark interval and its value of *p* for the next lowest benchmark interval. For example, if an item has a *p* of 0.7 at a benchmark (say benchmark 2) and a *p* of 0.5 at the next lower benchmark (say benchmark 1), its discrimination at benchmark 2 is 0.7-0.5 = 0.2. An item therefore would have the same number of discrimination values as the number of benchmarks. In this study, we will use 0.25 as the discrimination criterion for selecting exemplar items.

Method: Assuming the mean student proficiency for a CELLA scale is X and the standard deviation is Y, exemplar items will be selected for benchmark scores demarcated by the mean and standard deviation. That is, exemplar items will be identified for the following benchmark X–Y, X, X+Y and X+2Y on the vertical scale. The same benchmarks will be selected for Writing, Reading and Listening/Speaking scales. In order to obtain sufficient number of students for each benchmark, the benchmark intervals are defined as ¼ of a standard deviation above and below the benchmarks. The benchmark score intervals are defined as: (X–Y)±1/4Y, X±1/4Y, (X+Y)±1/4Y and (X+2Y)±1/4Y. The green highlighted area in Figure 1 demonstrates the four benchmark intervals in proficiency z-score scale. Notice that there is an extra highlighted area below benchmark level 1. This extra interval (below benchmark interval 1) is used to calculate the discriminations for items in benchmark interval 1. Note that alternative score intervals or benchmarks will be considered if sufficient numbers of exemplar items cannot be produced by the field test data.

Figure 1. Benchmark Intervals



We expect to have approximately 1000 students completing each item in the CELLA field test. This is an adequate number of students to apply the empirical midpoint method described in the Zwick et al. (2001) study for scale anchoring. The empirical method does not employ any assumption on population proficiency distribution as was done in some of the early anchoring NAEP work. The ESL population tends to be transient in nature so assuming a normal proficiency distribution might be misleading in generating exemplar items.

The RP (probability of correct response) *p* for an item under the empirical midpoint method is defined as the proportion of correct item responses for all individuals whose proficiency score falls within the benchmark intervals as indicated in Figure 1. For each MC item in the CELLA field test, a *p* value can be calculated for each of the benchmark intervals. For constructed-response items, *p* is calculated for each score level of the item. Different score levels of a given constructed-response item can potentially be exemplar items for more than one benchmark.

Items with *p* value equal to or greater than the RP criterion (RP-74 for multiple-choice items and RP-65 for constructed-response items) and discrimination > 0.25 will be selected as exemplar items for the given benchmark interval. Note that items being selected at a lower benchmark will not be eligible for other benchmarks. For example, if an item has a *p* value of 0.83 at benchmark 2 and a *p* value of 0.90 at benchmark 3, this item is eligible to be an exemplar item at benchmark 2 only. By definition, the

behavioral descriptions are hierarchical and students at benchmark 3 are likely to be able to answer items at benchmark 2 and 1.

After identification of the exemplar items for each benchmark, content experts will generate behavioral descriptions that characterize proficiency at the different benchmarks. These descriptions can be used to better inform stakeholders about what CELLA scores mean. Depending on how the field test data behave, we could explore alternative methods for the CELLA scale descriptors that describe what the CELLA vertical scale score means for narrower benchmark intervals instead of the broader intervals as proposed in this document.

Caveats: The scale anchoring procedures use arbitrary benchmarks to define benchmark intervals that are used in identification of exemplar items. These benchmarks used in scale anchoring do not carry the same meaning as they would if obtained in a formal standard setting, where panelists work from performance level descriptions ("A proficient student should be able to ….") to define what proficiency means in terms of student performance on the assessment. CELLA is an assessment that is intended for use in many states, and states can be different in their interpretation of the meaning of proficiency. This scale anchoring procedure adds behavioral meaning to the scale scores provided by CELLA and is not intended to replace states' individual formal standard settings, which will define student expectations in each state.

References

Beaton, A. E. & Allen, N. L. (1992) Interpreting scales through scale anchoring. Journal of Educational Statistics, 17, 191–204.

Zwick, R., Senturk, D., Wang, J. & Loomis, S. (2001). An Investigation of Alternative Methods for Item Mapping in the National Assessment of Educational Progress. Educational Measurement: Issues and Practice, 20(2), 18–25.

Appendix D: Directions for Using the Locator Test

Directions for Using the Locator Test

There are several important things to keep in mind when considering functional level testing with the Locator Test:

- Functional level testing only applies to the Reading and Writing sections. It does not apply to the Listening and Speaking sections. Students should take the Listening and Speaking sections of the test level corresponding to their actual grade levels.
- The Locator Test should only be given to students in grades 3 and above. Students in kindergarten through grade 2 should take the Level A test book of CELLA that contains the following sections: Listening, Reading, Writing, and One-on-One.
- The Locator Test can only indicate that a student should take the CELLA Reading and Writing sections at a level below that corresponding to the student's grade level. Students should never take any section of CELLA at a level higher than that corresponding to their grade level.
- The Locator Test is not a placement test and should not be used to place students into ESL or bilingual education classes. Its only purpose is to provide information about what level of the CELLA Reading and Writing sections is most appropriate for a student.

Test Administration

The Locator Test must be given far enough in advance of the scheduled date for the administration of the Reading and Writing sections to allow time for the Locator Test to be scored and for the appropriate level of the Reading and Writing sections for each student to be determined.

Total administration time for the Locator Test is approximately 35 minutes. This includes 5 minutes for distribution and collection of materials, 5 minutes for explanation of directions, and 25 minutes for students to read the four passages on the Locator Test and answer the 18 questions about them.

Scoring and Interpreting Locator Test Scores

The purpose of the CELLA Locator Test is to provide general information about a student's reading level. This information should be used along with teacher judgment, classroom performance, and all other available relevant information in deciding which level of the CELLA Reading and Writing sections a student should take.

There are three steps to scoring and interpreting the CELLA Locator Test:

- Determining the raw score
- Determining the recommended test level
- Deciding on the appropriate test level for the Reading and Writing sections

Each of these steps is explained below.

Determining the Raw Score. The Locator Test is designed to be hand scored locally (i.e., scored by a teacher or test administrator). The student's raw score is simply the number of items on the Locator Test that were answered correctly.

Determining the Recommended Test Level. The table below is used to determine the recommended test level for the student based on the student's raw score.

Raw Score on Locator Test	Recommended Test Level
0–5	Level A
6–8	Level B
9–12	Level C
13–18	Level D

Deciding on the Appropriate Test Level for the Reading and Writing Sections. The CELLA Locator Test provides one piece of information about which level of the Reading and Writing sections is most appropriate for the student. Each student should take the level of the Reading and Writing sections that will provide the best information about the student's English language proficiency, that will challenge, but not frustrate the student, and that will provide appropriate baseline data from which the student's progress can be measured.

The recommended test level is an indication of which test level of the CELLA Reading and Writing sections may be most appropriate for the student. However, other factors should also be considered in determining the actual test level of the CELLA Reading and Writing sections. These include:

- Students should not take a test level higher than that corresponding to their grade. For example, a student in grade 3 (Level B) should not take a Level C or Level D test, regardless of the score on the Locator Test.
- Teacher judgment and student classroom performance should be strongly considered in assigning a test level.
- The table below lists possible test levels that may be appropriate in an individual situation:

Student's Grade*	Possible Test Levels for Reading and Writing
3-5 (Level B)	Levels B or A
6-8 (Level C)	Levels C, B, or A
9-12 (Level D)	Levels D, C, B, or A

*Grades K–2 are not included in this chart because all students in kindergarten through grade 2 should take the Level A test book that includes the following sections: Listening, Reading, Writing, and One-on-One.

Appendix E: Field Test Demographics

Tests N in each modality by Home Language

Tested N	modalty			
HomeLang	**Oral**	**Read**	**Write**	**Grand Total***
**	98	106	106	310
Arabic	633	635	641	1,909
Chaldean	61	61	61	183
Chinese	407	408	415	1,230
French	200	201	207	608
Haitian Creole	434	448	450	1,332
Korean	350	348	360	1,058
NA	62	59	65	186
Other	2,522	2,552	2,585	7,659
Russian	116	118	120	354
Spanish	9,656	9,860	10,282	29,798
Urdu	145	145	148	438
Vietnamese	346	358	363	1,067
(blank)	61	66	66	193
Grand Total	15,091	15,365	15,869	46,325

Tested %	modalty			
HomeLang	Oral	Read	Write	Grand Total
**	0.6	0.7	0.7	0.7
Arabic	4.2	4.1	4.0	4.1
Chaldean	0.4	0.4	0.4	0.4
Chinese	2.7	2.7	2.6	2.7
French	1.3	1.3	1.3	1.3
Haitian Creole	2.9	2.9	2.8	2.9
Korean	2.3	2.3	2.3	2.3
NA	0.4	0.4	0.4	0.4
Other	16.7	16.6	16.3	16.5
Russian	0.8	0.8	0.8	0.8
Spanish	64.0	64.2	64.8	64.3
Urdu	1.0	0.9	0.9	0.9
Vietnamese	2.3	2.3	2.3	2.3
(blank)	0.4	0.4	0.4	0.4
Grand Total	100	100	100	100

* Total = # of tests, not # of students