Iowa Assessments Research and Development Guide

Form E and Form F All Levels

Contents

Validity

In Brief

Validity is an attribute of information from tests that, according to the *Standards for Educational and Psychological Testing*, "refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (1999, p. 9).

Assessment information is not considered valid or invalid in any absolute sense. Rather, the information is considered valid for a particular use or interpretation and invalid for another. The Standards further state that validation involves the accumulation of evidence to support the proposed score interpretations.

This part of this guide provides an overview of the data collected over the history of the *Iowa Assessments* that pertain to validity. Data and research pertaining to the *Iowa Assessments* consider the five major sources of validity evidence that are outlined in the Standards:

- test content
- response processes
- internal structure
- relations to other variables
- consequences of testing

The rationale for the professional judgments that lie behind the content standards and organization of the *Iowa Assessments* and the process used to translate those judgments into developmentally appropriate test materials are presented below. A range of appropriate uses of results and methods for reporting information on test performance to various audiences are also described.

Criteria for Evaluating Achievement Tests

Evaluating an elementary school achievement test is much like evaluating other instructional materials. In the latter case, the recommendations of other educators as well as of the authors and publishers would be considered. The decision to adopt materials locally, however, would require page-by-page scrutiny of the materials to understand their content and organization. The alignment of the materials with local educational standards and compatibility with instructional methods would be important factors in the review of the materials.

The evaluation of an elementary achievement test is much the same process. What the authors and publisher can say about how the test was developed, what the statistical data indicate about the technical characteristics of the test, and what judgments about quality made by unbiased experts as they review the test all contribute to the final evaluation. The decision about the potential validity of the test, however, rests primarily on local review and item-by-

item inspection of the test itself. Local analysis of test content—including judgments about its appropriateness for students, teachers, other school personnel, and the community at large—is critical.

Validity of the Tests

Validity must be judged in relation to purpose. Different purposes may call for tests built to different specifications. For example, a test intended to determine whether students have reached a performance standard in a local district is unlikely to have much validity for measuring differences in progress toward individually determined goals. Similarly, a testing program designed primarily to answer "accountability" questions may not be the best program to stimulate differential instruction and creative teaching.

Cronbach long ago made the point that validation is the task of the interpreter: "In the end, the responsibility for valid use of a test rests on the person who interprets it. The published research merely provides the interpreter with some facts and concepts. He has to combine these with his other knowledge about the person he tests...." (1971, p. 445). Messick contended that published research should bolster facts and concepts with "some exposition of the critical value contents in which the facts are embedded and with provisional accounting of the potential social consequences of alternative test uses" (1989, p. 88). More recently, Kane proposed that validation is a way of thinking about the use of test results that (1) establishes a framework for test development based in the interpretations to be made of test results, (2) structures the evidence that should be gathered to support an argument for validity of the intended interpretations, and (3) clarifies the extent to which the argument for validity is adequate for the purpose the test is intended to serve (2006, p. 60). All of these perspectives reflect important aspects of validity in large-scale assessment.

Instructional decisions involve the combination of test validity evidence and prior information about the person or group tested. The information that test developers can reasonably be expected to provide about all potential uses of tests in decision-making is limited. Nevertheless, one should explain how tests are developed and provide recommendations for appropriate uses. In addition, guidelines should be established for reporting test results that lead to valid score interpretations so that the consequences of test use at the local level are clear.

The procedures used to develop and revise test materials and interpretive information lay the foundation for test validity. Meaningful evidence related to inferences based on test scores, not to mention desirable consequences from those inferences, can provide test scores with social utility only if test development produces meaningful test materials. Content quality is thus the essence of arguments for test validity (Linn, Baker & Dunbar, 1991). The guiding principle for the development of the *Iowa Assessments* is that materials presented to students be of sufficient quality to make the time spent testing instructionally useful. Passages are selected for the reading tests, for example, not only because they yield good comprehension questions, but because they are interesting to read. Items that measure discrete skills (for example, capitalization and punctuation) contain factual content that promotes incidental learning during the test. Experimental contexts in science expose students to novel situations through which their understanding of scientific reasoning can be measured. These examples

show ways in which developers of the *Iowa Assessments* try to design tests so that taking the test can itself be considered an instructional activity. Such efforts represent the cornerstone of test validity.

Statistical Data to Be Considered

The types of statistical data that might be considered as evidence of test validity include reliability coefficients, difficulty indices of individual test items, indices of the discriminating power of the items, indices of differential functioning of the items, and correlations with other measures such as course grades, scores on other tests of the same type, or experimental measures of the same content or skills.

All of these types of evidence reflect on the validity of the test, but they do not guarantee its validity. They do not prove that the test measures what it purports to measure. They certainly cannot reveal whether the things being measured are those that ought to be measured. A high reliability coefficient, for example, shows that the test is measuring something consistently but does not indicate what that "something" is. Given two tests with the same title, the one with the higher reliability may actually be the less valid for a particular purpose (Feldt, 1997). For example, one can build a highly reliable mathematics test by including only simple computation items, but this would not be a valid test of problem-solving skills. Similarly, a poor test may show the same distribution of item difficulties as a good test, or it may show a higher average index of discrimination than a more valid test.

Correlations of test scores with other measures are evidence of the validity of a test only if the other measures are as good as or better than the test that is being evaluated. Suppose, for example, that three math tests, A, B, and C, show high correlations among themselves. These correlations may be due simply to the three tests exhibiting the same defects, such as overemphasis on memorization of basic facts. If Test D, on the other hand, is a superior measure of the student's ability to apply those math principles to real-world problems, it is unlikely to correlate highly with the other three tests. In this case, its lack of correlation with Tests A, B, and C is evidence that Test D is the more valid test for interpretations about problem solving.

This is not meant to imply that well-designed validation studies are of no value; published tests should be supported by a continuous program of research and evaluation. Rational judgment also plays a key part in evaluating the validity of achievement tests against content and process standards and in interpreting statistical evidence from validity studies.

Validity of the Tests in the Local School

Standardized tests such as the *Iowa Assessments* are constructed to correspond to widely accepted goals of instruction in schools across the nation. No standardized test, no matter how carefully planned and constructed, can ever be equally suited for use in all schools. Local differences in curricular standards, grade placement, and instructional emphasis, as well as differences in the nature and characteristics of the student population, should be taken into account in evaluating the validity of a test.

The two most important questions in the selection and evaluation of achievement tests at the local level should be:

1. Are the skills and abilities required for successful test performance those that are appropriate for the students in our school?
2. Are our standards for content and instructional practices represented in the test questions?

To answer these questions, those making the determination should take the test or at least answer a sample of representative questions. In taking the test, they should try to decide by which cognitive processes the student is likely to use to reach the correct answer. They should then ask:

- Are all the cognitive processes considered important in the school represented in the test?
- Are any desirable cognitive processes omitted?
- Are any specific skills or abilities required for successful test performance unrelated to the goals of instruction?

Evaluating an achievement test battery in this manner is time-consuming. It is, however, the only way to discern the most important differences among tests and their relationships to local curriculum standards. Considering the importance of the inferences that will later be drawn from test results and the influence the test may exert on instruction and guidance in the school, this type of careful review is important.

Domain Specifications

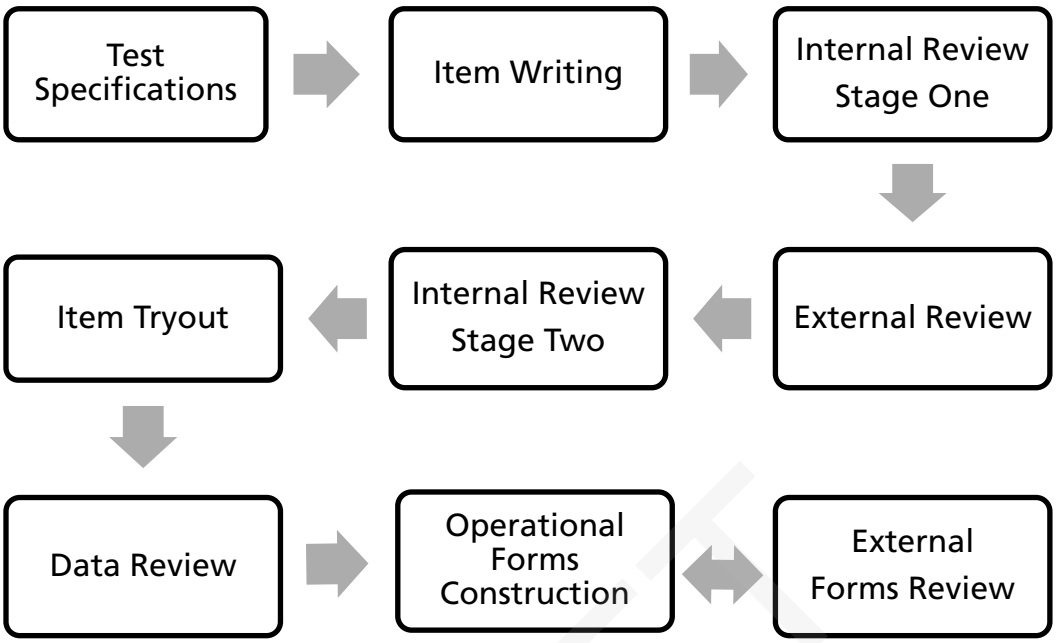
The content and process specifications for the *Iowa Assessments* have undergone constant revision for more than 60 years. They have involved the experience, research, and expertise of professionals from a variety of educational specialties. In particular, research in curriculum practices, test design, technical measurement procedures, and test interpretation and utilization has been a continuing feature of test development.

Forms E and F of the *Iowa Assessments* reflect today's curricula: the tests have been carefully designed using the Common Core State Standards (CCSS), individual state standards, surveys of classroom teachers, reviews of curriculum guides and instructional materials, and responses from students in extensive research studies and field testing.

Test Development Procedures

The new forms of the *Iowa Assessments* are the result of an extended, iterative process during which "experimental" test materials are developed and administered to national and state samples to evaluate their measurement quality and appropriateness. Figure 1 shows the process involved in test development.

Figure 1: Steps in Development of the *Iowa Assessments*



Test Specifications

Test specifications are created that outline (among other attributes) the statistical specifications; distribution of content, skills, and cognitive levels across the test form; test organization; and special accommodations and other conditions of test administration. By establishing these parameters beforehand, test specifications also help to ensure the new forms are comparable to existing forms to the degree desired. The test specifications provide the "blueprint" for test construction, defining the necessary steps and procedures. As test development proceeds, the test specifications are continually revisited and evaluated in an iterative process to ensure that the materials available for assembly of final forms reflect the evolving purposes of the assessments.

Item Writing

Items and stimulus/item sets (reading passages, graphs, maps, tables, and so on that support a group of items) are then created according to the test specifications. Content specialists at ITP convene item writing workshops and train educators on sound item writing practices. Educators are assigned to write items in the content areas and grade levels that best align with their experience in the classroom. Item production goals ensure a significant "overage" of items across subject areas at each cognitive level so that the pool of available items in each subject and level is far greater than is needed to build each test. This overage allows content experts to discard those items that do not survive internal and external item review or post-tryout data review.

Internal Review Stage One

After items are written, content specialists review these items for content accuracy, fairness, and universal design (see "Universal Design" on page 29 for more information). The goal of these reviews is to make sure the items are accurate, fair, and accessible to all student subgroups in the diverse population of test takers. The items and associated materials are edited to ensure that they are clearly written and that reading loads are grade appropriate. The items are also copyedited for grammar and spelling at this stage in the process.

External Review

Once the items have been reviewed internally, ITP convenes panels of educators to review the items and associated stimuli (reading passages, tables, graphs, maps, and so forth). After participating in a formal training session about the review process, educators review the items for grade-level appropriateness, content relevance, and accuracy. Since they have not been involved in the development process up to this point, external reviewers provide an objective "cold read" of potential test materials. A main goal of the educator review is to confirm that the items are appropriate for the intended grade level and content area.

Internal Review Stage Two

ITP development staff reviews the items again after the educator panel review. This review focuses on edits made to the items during previous steps in the process and again checks for content accuracy, fairness, and universal design considerations.

Item Tryout

Items that have passed the review process are assembled into field test forms for the Item Tryout. ITP collects data on the performance of the items by conducting a field test to determine how well the items are likely to perform operationally. When a field test is conducted, test booklets are created to be tried out at predetermined grade bands spanning 2, 3, or 4 grade levels. Students complete the field tests when they take the operational tests in numbers sufficient to ensure the associated statistical results are sound. Trying out test materials at multiple grades provides the data necessary to ensure optimal placement of items for the measurement of growth.

Data Review

The data collected during the field test are analyzed for technical qualities related to item difficulty and discrimination. This analysis determines whether the items are appropriate measures of students' knowledge and the extent to which they will contribute to the test's overall reliability. Other aspects of the Data Review include key checks and the analysis of distractor choices, subgroup differences, and correlations with operational test forms. Only items that display acceptable descriptive statistics are eligible to appear on operational forms.

Operational Forms Construction

Items that ITP has determined should appear on operational test forms become part of the pool of items that are eligible for selection. Forms construction procedures ensure the final subject area test has adequate content coverage while being meaningful to students of varying achievement levels, the items within a typical subject area's item pool are diverse in

terms of skill alignment, cognitive level, and difficulty. Items are then selected from the item pool into test forms. Careful attention is paid to item selection so that the final tests follow the predetermined test specifications and meet psychometric targets for difficulty, discrimination, and reliability.

Forms Review

Once tests have been constructed, the materials are submitted for another round of external reviews. Educators are recruited to evaluate the materials from a variety of perspectives including appropriateness for the intended audience. Additionally, experts are recruited to evaluate materials for perceived fairness and sensitivity concerns. Educators/reviewers are selected to represent various ethnic and racial groups, genders, and student subgroups such as English Language Learners (ELLs), students with special needs, and students who are visually impaired (the latter aids in the adaptation of test forms in braille).

Test Descriptions

The following tables provide a description of each subject-area test in the *Iowa Assessments*, grouped by level as appropriate. As students progress through the elementary, middle, and high school grades and gain greater mastery in a given subject area, the skills and concepts on which they are assessed change accordingly. Broadly speaking, each assessment can be viewed as measuring a continuum of achievement that spans ages 5/6 through 18, which are referred to as test levels.

Level 5/6

Test	Description
Vocabulary	• Questions measure listening vocabulary • Students hear a word and select a picture that illustrates the meaning of the word • Nouns, verbs, and modifiers are included
Word Analysis	• Questions emphasize the recognition of letters and letter-sound relationships • Response choices are a mix of letters, pictures, or words
Listening	• Questions emphasize literal and inferential understanding of material that is heard • Stories are read aloud and followed by a question • Response choices are pictorial • Reading is not required

Continued on the next page…

Level 5/6, *continued*

Test	Description
Language	• Questions measure the student's ability to use language to express ideas • Some questions cover the use of prepositions, singular and plural, and comparative and superlative forms • Some questions are aimed at word classifications, verb tenses, or spatial-directional relationships • Questions are read aloud • Response choices are pictorial
Mathematics	• Questions emphasize beginning math concepts, problem solving, and math operations • Questions are drawn from numeration, geometry, measurement, and applications of addition and subtraction in word problems • Questions are read aloud • Response choices are pictorial and numeral
Reading	• Administered in two parts • Questions emphasize the ability to identify words based on verbal and visual cues • Measures comprehension of sentences, pictures that tell a story, and printed stories

Levels 7 and 8

Test	Description
Vocabulary	• Students are presented with a pictorial or written stimulus and select the answer from a set of written responses • Nouns, verbs, and modifiers are included • Content focus is on general vocabulary • Test consists of two untimed sections
Word Analysis	• Questions measure comprehension of letter-sound associations and word structures using affixes and the formation of compound words • Response choices are a mix of pictures and words
Reading	• Administered in two parts • Questions emphasize the ability to complete sentences based on visual cues • Measures the ability to demonstrate both literal and inferential understanding
Listening	• Questions emphasize literal and inferential understanding of material that is heard • Stories are read aloud and followed by one or more questions • Response choices are pictorial

Continued on the next page...

Levels 7 and 8, *continued*

Test	Description
Language	• Questions measure the student's ability to use some conventions of standard written English • Four test sections assess spelling, capitalization, punctuation, and skill in written usage and expression • Questions and response choices are read aloud
Mathematics	• Administered in two untimed parts • Questions measure the understanding and ability to apply concepts in the areas of number properties and operations, geometry, measurement, and number sentences • Questions emphasize the interpretation of data presented in graphs or tables, where students response options are either pictorial, numbers, or words • Some questions require students to select a number sentence that could be used to solve the problem, while other questions require students to solve brief word problems with answer options that include "N," indicating that the solution is not provided with the answer choices • If the correct answer is not given, students select "N," which means "Not given" • Questions are read aloud
Computation	• First section is an oral presentation of addition and subtraction problems • Second section is not read aloud and addition and subtraction questions are presented in the test booklet • If the correct answer is not given, students select "N," which means "Not given"
Social Studies	• Questions emphasize the interpretation of social studies-related materials, as well as knowledge drawn from the areas of history, geography, economics, civics, and government • Most questions are read aloud • Response choices are pictorial or text • At the end of the test, students respond to sets of stimuli (questions and stimuli are not read aloud)
Science	• Questions emphasize the methods and processes used in scientific inquiry, as well as knowledge in the areas of life science, earth and space science, and physical science • Most questions are read aloud • Response choices are pictorial or text • At the end of the test, students respond to sets of stimuli (questions and stimuli are not read aloud)

Levels 9–14

Test	Description
Reading	• Administered in two parts • Includes both literary and informational passages • Questions focus on identifying, interpreting, analyzing, and extending information in passages
Written Expression	• Some questions focus on the most appropriate way to express the ideas in a piece of writing • Some questions focus on the identification of the line of text that contains an error • Questions may address organization, sentence structure, clarity, and effective or inappropriate language
Mathematics	• Administered in two parts • Questions are drawn from the areas of number sense and operations, algebraic patterns and connections, data analysis/probability/statistics, geometry, and measurement
Science	Questions emphasize the methods and processes used in scientific inquiry, as well as knowledge in the areas of life science, earth and space science, and physical science
Social Studies	Questions emphasize the use and understanding of concepts, principles, and various types of visual materials such as posters, cartoons, timelines, maps, graphs, tables, charts, and passages
Vocabulary	• Questions emphasize general vocabulary words in the context of a short phrase or sentence • Students select the answer that is the closest synonym for the given word • Includes nouns, verbs, and modifiers
Spelling	• Questions emphasize errors in root words, such as substitutions, reversals, omissions, and errors associated with suffixes • Each question presents four words, one of which may be misspelled, and a fifth option, "No mistakes"
Capitalization	• Questions emphasize errors in the capitalization (under punctuation and over punctuation) of names, dates, and other words • Students mark the line of text that contains an error • If there is no error, students select "No mistakes"
Punctuation	• Questions emphasize errors in the use in punctuation (under punctuation and over punctuation), commas, and other punctuation such as quotation marks • Students mark the line of text that contains an error • If there is no error, students select "No mistakes"
Computation	• Questions emphasize addition, subtraction, multiplication, or division using whole numbers, fractions, or decimals • In Level 14, some questions emphasize algebraic manipulation • If the correct answer is not given, students select "N" which means "Not given"

Levels 15–17/18

Test	Description
Reading	• Questions measure the ability to understand a range of process levels associated with reading comprehension • Each test level has five passages • Questions focus on inferring, analyzing, evaluating, and generalizing information in passages
Written Expression	• Questions measure the ability to recognize the correct and effective use of standard American English in writing • Some questions focus on the most appropriate way to revise a piece of writing based on focus, organization, diction and clarity, sentence structure, usage, mechanics, and spelling • Questions pose alternatives that may correct or improve underlined portions of texts, including: errors in mechanics or usage, problems with fluency or clarity, or issues of organization
Mathematics	• Questions measure the students' ability to solve quantitative problems • Problems require basic arithmetic and measurement, estimation, data interpretation • Questions are drawn from the areas of number sense and operations, algebraic patterns and connections, data analysis/probability/statistics, geometry, and measurement
Science	• Questions emphasize the methods and processes used in scientific inquiry • Questions assess knowledge and skill in life science, earth and space sciences, and physical science
Social Studies	• Questions emphasize the use and understanding of concepts, principles, and various types of visual materials such as posters, cartoons, timelines, maps, graphs, tables, charts, and passages • Questions are drawn from knowledge in the areas of history, geography, economics, and civics and government
Vocabulary	• Questions represent a cross section of vocabulary in general communication • Technical words and specialized vocabulary are not included • Words are presented in short sentences, and the student must chose an alternative word or phrase that is closest in meaning to the tested word
Computation	• Questions emphasize addition, subtraction, multiplication, and division using whole numbers, fractions, decimals, and percentages • Questions measure the ability to manipulate variables and to evaluate expressions with exponents or with square roots

Distribution of Domains and Skills for the *Iowa Assessments*

Table 1 lists the distribution of domains and skills in Levels 5/6 through 17/18 of the *Iowa Assessments*. The table indicates major categories in the test specifications for each test during item development.

Table 1: Distribution of Skills Objectives for the *Iowa Assessments*, Form E

Test	Level 5/6		Levels 7 and 8		Levels 9–14		Levels 15–17/18	
	Number of Domain Skills	Number of Standards	Number of Domain Skills	Number of Standards	Number of Domain Skills	Number of Standards	Number of Domain Skills	Number of Standards
Reading	2	6	3–5	8–11	5	11–12	5	11–12
Math	4	10	5	13–15	5	20–22	5	16–19
Written Expression	–	–	–	–	4	14–16	5	17–18
Science	–	–	3	11–12	3	10–11	3	10–12
Social Studies	–	–	4	10	4	9–12	4	10–11
Vocabulary	1	3	1	3	1	3	1	3
Computation	–	–	1	4	1–4	7–19	4	8
Spelling	–	–	–	–	1	5	–	–
Capitalization	–	–	–	–	1	9–12	–	–
Punctuation	–	–	–	–	1	6–9	–	–
Word Analysis	2	6	2	7	2*	8*	–	–
Listening	2	8	2	8	2*	8*	–	–
Language	7	7	4	14–15	–	–	–	–
Common Core Reading	–	–	3**	–	3	–	3	–
Common Core Foundational Skills	–	–	2**	–	2***	–	–	–
Common Core Speaking and Listening	–	–	1**	–	1***	–	–	–
Common Core Language and Writing	–	–	2**	–	5	–	5	–
Common Core Mathematics	–	–	4**	–	5–6	–	5	–

*Word Analysis and Listening are supplementary tests at Level 9.
**Level 8 only
***Level 9 only

Cognitive Level Difficulty Descriptors

In order to help educators see the full range of item complexity in the *Iowa Assessments* and how their students perform on items of varying cognitive complexity, each item in Forms E and F has been assigned one of three Cognitive Level Difficulty descriptors:

Level 1: Essential Competencies	This level of difficulty involves recalling information such as facts, definitions, terms, or simple one-step procedures.
Level 2: Conceptual Understanding	This level of difficulty requires engaging in some cognitive processing beyond recalling or reproducing a response. A conceptual understanding item requires students to make some decisions as to how to approach the problem or activity and may require them to employ more than a single step.
Level 3: Extended Reasoning	This level of difficulty requires problem solving, planning, and/or using evidence. These items require students to develop a strategy to connect and relate ideas in order to solve the problem, and the problem may require that the student use multiple steps and draw upon a variety of skills.

Internal Structure of the *Iowa Assessments*

The internal structure of the *Iowa Assessments* was analyzed using exploratory factor analysis (EFA) techniques. In general, the results of these analyses, particularly in grades 3 through 11, reflect a composition of constructs consistent with the major domains of the Common Core State Standards: (1) reading and writing aspects of literacy in connection with analysis of information in social studies and science, and (2) concepts and procedural skills in mathematics. Correlations among national standard scores were used with least-squares estimates of communality.

- In Kindergarten, the factor solution was based on the six tests in Level 5/6.
- In grades 1 and 2, the factor solutions were based on the nine tests in Levels 7 and 8.
- In grades 3–8, the solutions were based on the twelve tests in Level 9 and the ten tests in Levels 10–14.

After the least-squares factor solutions were obtained, both orthogonal and oblique simple structure transformations were performed. Three factors were retained for Levels 7–14; two were retained for Level 5/6.

For Levels 9–14, the three factors were primarily determined by tests in Reading, Vocabulary, Reading, Written Expression, and Mathematics. These tests are considered the major subject areas in the elementary school curriculum and are consistent with the emphasis found in the CCSS. Tests in Social Studies, Science, and Spelling were less uniform in their factor composition and loaded on a secondary factor.

The three constructs were identified as a "literacy" factor, a "mathematics" factor, and a "mechanics of written language" factor. The "literacy" factor was determined by subtests under the umbrella of English Language Arts. Among those, Vocabulary and Reading

contributed the most to the interpretation of this factor, with substantial influence from Written Expression, Social Studies, and Science. The inclusion of the Social Studies and Science tests in the literacy factor is also consistent with the structure of the CCSS, which includes processes such as the following in the ELA Literacy standards:

- Using textual evidence to support analysis of science and technical texts
- Determining the central ideas or information of a primary or secondary source (or of a text)
- Using textual evidence to support analysis of primary and secondary sources, connecting insights gained from specific details

The Mathematics test and the Computation test clearly identified the "mathematics" factor. The factor was the most clearly defined of the three. Although it was correlated with the other achievement constructs, it suggests that the mathematics domain in the *Iowa Assessments* is focused on well-defined and coherent standards of the curriculum.

The "mechanics of written language" factor was principally defined by subtests in Conventions of Writing (Spelling, Capitalization, and Punctuation). Written Expression also loaded on this factor at six test levels. This could be expected because the Written Expression test contains questions about specific points of Standard English syntax, verb forms, and other points of grammar often taught in conjunction with all aspects of written language and used by students editing their own written work or the work of their peers.

Levels 7 and 8 have a subtest structure similar to that of Levels 9–14 except in ELA, where the test specifications have features unique to these test levels. The three factors defined reveal contrasts between the tests in Level 7 and 8 and those in Levels 9–14. The first two factors were similar to the ones described above. The Word Analysis and Language tests helped define the first factor; the two Mathematics tests defined the second factor. The third factor related to the tests that require interpreting stories and pictures (Listening, Social Studies, and Science) while listening to a teacher as they are read aloud.

Only six tests are included in Level 5/6, and the test composition is slightly different from that of the higher levels. Two factors were defined at this level. The first factor was defined by the Vocabulary, Listening, Language, and Mathematics tests. The second factor was influenced by the Reading, Word Analysis and, to a lesser extent, Mathematics tests. The two factors probably reflect the integrated curriculum of the early elementary grades.

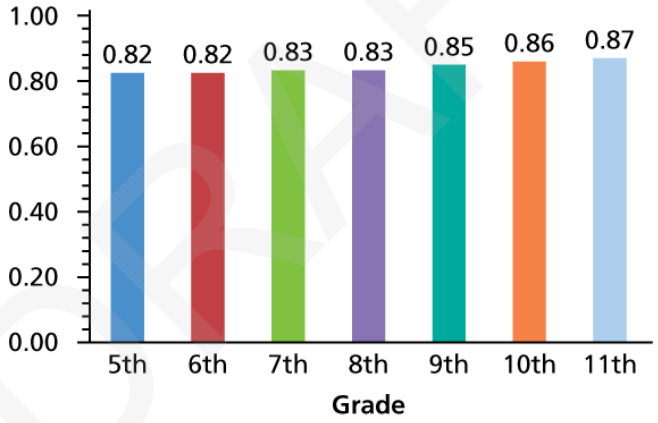
A similar procedure was performed at the high school level. The factor solutions were based on correlations among the seven tests for Levels 15 through 17. Again, three factors were retained for all three levels. The "literacy" factor was determined by the Reading, Vocabulary, and Written Expression tests. The "mathematics" factor was defined by the Mathematics and Computation tests. The third factor was primarily defined by Social Studies and Science, which require analysis of a variety of stimulus materials and questions tapping broad reasoning skills and principles of interpreting results of empirical research in science and social science. The Vocabulary test also loaded in this factor.

Predictive Validity and College Readiness

Tests such as the *Iowa Assessments* have been used in many ways to support judgments about how well students are prepared for future instruction—that is, as general measures of readiness. Over the years, ITP has conducted numerous studies to establish the predictive "power" of the *Iowa Assessments* with respect to a variety of criterion measures, including high school GPA, college GPA, and scores on college entrance exams such as the ACT® and SAT® (for example, Scannell, 1958; Rosemeier, 1962; Loyd, Forsyth, and Hoover, 1980, Ansley & Forsyth, 1983; Iowa Testing Programs, 1999). The *Guide for Research and Development, Forms A and B* includes the details of these studies.

More recently, Furgol, Fina, and Welch (2011) investigated the relationship between performance on the *Iowa Assessments* and college admissions test scores in a matched longitudinal cohort of more than 25,000 students in grades 5 through 11 who tested annually over a five-year period. Evidence of a strong relationship between *Iowa Assessments* scores and the ACT composite score suggests that the *Iowa Assessments* and college readiness measures assess the same achievement domains. As shown in Figure 2, this relationship sustains itself and strengthens from grades 5 to grade 11.

Figure 2: Correlations between *Iowa Assessments* and ACT composite scores



Furgol, Fina, and Welch (2011) also reported the correlations between ACT and *Iowa Assessments* subject-area test scores for approximately 18,000 students in grades 8–11. The correlations are reported in Table 2.

Table 2: Correlations between ACT and *Iowa Assessments* Content Area Test Scores

Grade	Reading	English	Math	Science
8	.74	.72	.75	.60
9	.75	.76	.74	.65
10	.72	.79	.75	.67
11	.75	.76	.76	.68

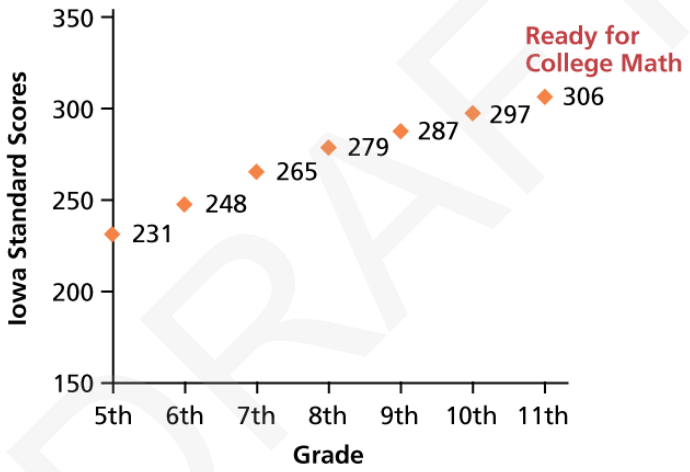
Each correlation in the table is based on the students who have both an ACT score in the subject area of interest and an *Iowa Assessments* score in both the subject area and grade of interest. These correlations are generally highest in grade 11, ranging from .68 (Science) to .76 (English and Math), providing supporting evidence for the use of the grade 11 *Iowa* scores to

predict whether students are likely to meet or exceed the ACT College-Readiness Benchmarks. Note that the unadjusted correlations between the grade 11 *Iowa Assessments* subject area tests and the corresponding ACT tests are as high as or higher than those between corresponding subject area tests on EXPLORE® and ACT, which are .68 for Reading, .75 for English, .73 for Math, and .65 for Science.

Tracking Readiness for Post-Secondary Education

In addition to the results described above, Furgol, Fina, and Welch (2011) linked the scores of grade 11 examinees on four *Iowa Assessments* subject-area tests to defined targets of readiness based on ACT scores. The linking method was based on the principle of balancing false positive and false negative probabilities in determining whether or not a student was likely to exceed or fall short of the ACT readiness benchmark. Once this link was established, the study then used the NSS scale of the *Iowa Assessments* to establish an on-track projection of college readiness for middle- and high-school grades, as illustrated by the example in Figure 3 for Mathematics.

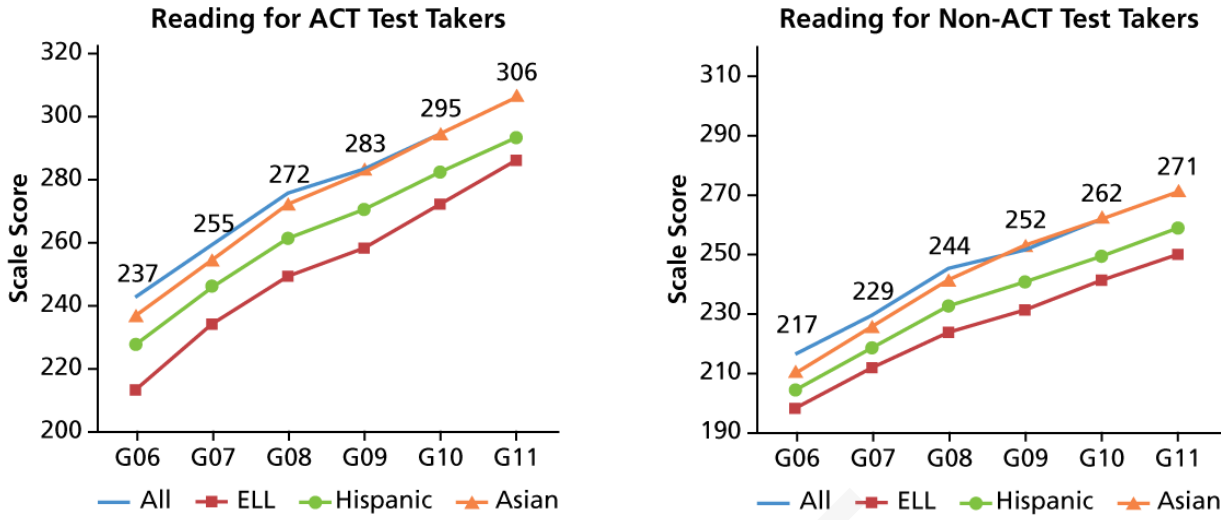
Figure 3: On Track to College Readiness in Mathematics



The results of ITP's research into the link between the *Iowa Assessments* and established college readiness benchmarks permits examinees to receive information on score reports that designate whether they are "On Track" or "Not Yet on Track" to be prepared for the first year of college in Reading, Language, Mathematics, and Science. In Figure 4, the "On Track" benchmarks scores on the NSS scale are marked. Examples of college readiness reports are included in the *Iowa Assessments Score Interpretation Guide, Levels 9–14* and in the *Iowa Assessments Score Interpretation Guide, Levels 15–17/18*.

A subsequent study by Wang, Chen, and Welch (2011) examined group differences in the empirical trajectories of performance and established that growth trends for culturally (for example, Asian and Hispanic) and linguistically diverse (that is, English Language Learners) test takers run parallel to the college readiness trajectories identified by Fina, et al. (2011). All effect sizes for departure from parallel trajectories were extremely small, as suggested by the results shown in Figure 4. Such results provide evidence of the appropriateness of using the NSS scale to track the college readiness of all students, in view of the subgroups included in this study.

Figure 4: On Track to College Readiness in Mathematics



Interpretation and Utility of Readiness Information

College readiness information gives educators and families information they need to determine whether students are on track to successfully complete first-year college coursework upon graduation from high school or whether additional coursework and preparation are necessary. It allows families and educators to monitor student progress from middle school through high school and allows flexibility to determine the appropriate improvement and support strategies for students as they plan for post-secondary education opportunities. Monitoring the use of readiness information of the type described here is an important responsibility at the local level. This information should be used in ways that inform instruction and enhance learning for students as they prepare for postsecondary education opportunities.

Validity in the Assessment of Growth

Score interpretations that provide for the assessment of student growth over time are an important aspect of large-scale assessment in education. The measurement of growth through the *Iowa Assessments* is based on the Iowa Growth Model and the underlying vertical scale used in reporting, the national standard score (NSS) scale. Vertical scaling is the term used for the process of linking assessments to describe student growth over time. Although the methods can be complex, the goal is quite simple: to create a framework and metric for reporting the educational development of individuals and groups. The challenge of vertical scaling of assessments has existed since the first use of standards-based assessments to measure individual and group progress (Patz, 2007) and it has a history that predates that work. Today, vertical scaling is needed for assessments of growth toward college and career readiness standards and for adaptive testing. In these applications, comparative information about results from assessments of different levels of difficulty is needed to build a vertical scale.

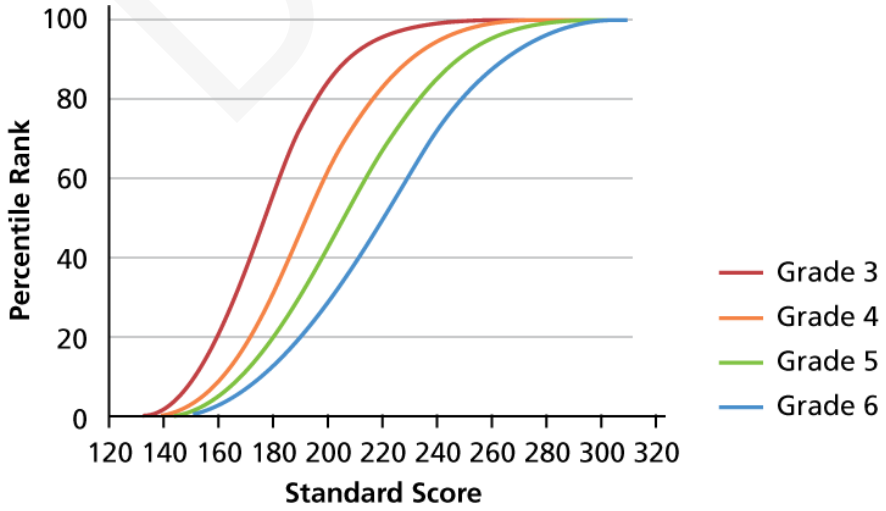
Assembling test forms with an evidence-based approach to growth on established content standards is a key element in vertical scaling. The methods used to build a vertical scale will

only work as intended if the assessment being scaled yields meaningful and stable changes in measured achievement across time. Assessments that are matched to content that are not vertically aligned across grade or that reflect an overly granular approach to domain definitions and content specifications may show irregular patterns of growth across grades for both individuals and groups.

The conceptual framework for a vertical scale is established when the content standards and learning progressions of the achievement domain are determined. In developing vertical scales for the *Iowa Assessments*, special assessments were designed in each content area based on prevailing sets of standards. These assessments were wide-range achievement tests consisting of items that spanned multiple grade levels to provide comparative information about the expected performance of students at different developmental levels of the content-area learning continuum. Comparative results on the special assessments administered across grades were used to define the range of student performance within each grade level and the amount of overlap between the distributions of student scores at different grade levels. Finally, a numerical scale that described the growth pattern observed on the special assessments was determined, resulting in the NSS scale.

The growth model for mathematics in grades 3 through 6 is illustrated in Figure 5. In the figure, each unit on the NSS scale is associated with a specific achievement level in each grade. From the plot, one can determine, for example, the percent of students scoring at or below an NSS of 185 in grade 3 (about 75) and that the equivalent achievement level in grade 4 corresponds to an NSS of about 204. Starting at 185 and following the arrows, the model says a student in grade 3 who scores 185 in math is expected to score 204 in grade 4, other things being equal. Similar relationships between NSS scores at multiple grade levels provides a comprehensive framework for determining growth expectations, comparing those expectations to observed growth and describing the value added or response to intervention resulting from students' instructional programs and learning experiences.

Figure 5: Standard Score Growth Model of the Iowa Assessments–The NSS Scale



Description and Primary Interpretation of the NSS Scale

The NSS scale is a metric that ranges numerically from 80 to 400 and spans a developmental continuum from kindergarten to grade 12 in major content domains such as reading, mathematics, science, social studies, and written expression.

National research studies in the 2010–11 school year were conducted to validate the reference points on the NSS scale that represent the medians for each grade level and the model-based inferences about the amount of growth typical of students at different achievement levels. The primary interpretations supported by the NSS scale have to do with (1) how much a student is growing from one assessment occasion to the next compared to his or her assessment peers (a relative growth interpretation), and (2) how much growth would be expected for this student's assessment peers (a normative growth interpretation). This basic information about growth can be used for a variety of purposes in student and program evaluation such as individual and group decisions about instructional interventions, and responses to interventions that can be gauged by the amount of growth achieved.

Validity Framework and Statistical Foundation of Growth Metrics

The validity framework for a growth model involves fundamental considerations about the content of the assessments used to measure growth, the scale and modeling requirements, the definition of targets that represent typical grade-level performance or other benchmarks such as college readiness, and the utility of information leading to sound interpretations of student growth and effective decisions about enhancing growth for individuals and groups.

Validity

In the context of achievement over time, validity pertains to evidence that supports interpretations relative to growth. With the assessment imperative of college- and career-readiness at the forefront of efforts to reform education, a critical aspect of validity arguments for related claims involves the underlying model used to measure and report growth and change. Psychometric frameworks for quantifying growth are evolving rapidly (e.g. Castellano & Ho, 2013; Betebenner, 2010; Reardon & Raudenbush, 2009). For any growth model, validity considerations encompass evidence that ranges from the content definition of the domain to the utility of growth reports. Regardless of the approach to growth, general validation concerns remain. Table 3 summarizes several of these issues as they define a validity framework for growth.

Table 3: Examples of Validity Evidence Related to the Measurement of Growth

Validity Evidence	Consideration for Growth
Content validity evidence	Content-related validity evidence is tied to test development. The proposed interpretations of growth and readiness should guide the development of a test and the inferences leading from the test scores to conclusions about a student's readiness. Content alignment studies will serve as the foundation for a trail of evidence needed for establishing the validity of growth and readiness tracking and reporting. Alignment studies will inform the interpretation of growth and readiness research findings from the statistical relationship studies and shape assessments that are making the claim to identify students who are on track.
Scale requirements	Scales or linking studies that allow for the longitudinal estimation and tracking of growth are a necessity in the present context. The scales need to be anchored in terms of both content and student performance within and between grades.
Definition of targets	Targets must exist that quantify the level of growth expected, observed and desired for a given period of time (that is, fall to spring testing; year to year testing). For college readiness, targets must also exist that quantify the level of achievement where a student is ready to enroll and succeed in credit-bearing, first-year postsecondary courses. To date, these targets are currently defined by the ACT Benchmarks, by the College Board Readiness Index, or by individual institutions of higher education.
Collection of concurrent validity evidence	Many tests will claim to measure college readiness, but a plan must be in place for validating that claim. Validity studies should be conducted to determine the relationship between the assessments and the indicators of readiness, including the content of entry-level college courses.
Utility	A primary goal of this information is that students, K-12 educators, policy makers and higher education representatives can use it to better understand the knowledge and skills necessary for college readiness in English language arts and mathematics. The information must be easily understand and actionable by a broad range of audiences.

Developing a domain and model for growth begins with defining content standards that describe *continuous* learning. Discrete, granular descriptions of content that are the objectives of small instructional units in, for example, signed-number arithmetic, may be useful in tracking progress toward small unit objectives, but they may not be the best focus for an assessment of growth being used to track progress across large spans of time, such as grade-to-grade growth over the elementary school years. The five stages of development in reading (Chall, 1996) are a good example of a learning continuum—there is an underlying construct and a progression that describes how children change from "learning to read" to "reading to learn." In this sense, the learning continuum constitutes a broad definition of the achievement domain and what it means to "grow" with respect to important content standards or guideposts of the domain. The important point is that measuring growth requires test design and development that keeps the focus on the domain.

Assessing a child's growth on a learning continuum requires measures aligned to broad content standards and a level of cognitive complexity appropriate for that child's stage of

development. Developmental appropriateness is (1) guided by research and practice in the achievement domain (for example, the major domains of the Common Core State Standards in English language arts), and (2) established through extensive field testing of assessment materials *in multiple grades*. Valid and reliable measurement of growth requires both.

Statistical Foundation

The NSS scale of the *Iowa Assessments* quantifies and describes student growth over time via a growth metric. One of the defining attributes of the growth metric is the projection of subsequent performance can be made conditional on prior performance through the vertical scale (Furgol, Fina & Welch, 2011). The expected NSS scores for each grade level and content area on the *Iowa Assessments* show the relative standing of students' achievement within the score distribution of students in a national probability sample (Hoover, et al., 2007).

Many tests used to measure yearly growth are vertically aligned and scaled. This means that each successive test builds upon the content and skills measured by the previous test. It assures that tests taken over multiple grade levels show a coherent progression in learning. They incorporate a several defining technical characteristics (Patz, 2007), including:

- an increase in difficulty of associated assessments across grades
- an increase in scale score means with grade level, and
- a pattern of increase that is regular and not erratic

Assessing annually does not necessarily mean that the change in scores reflects a year's growth in student achievement. That is where vertical scaling comes in: Tests are developed for different grade levels—for example, for grades 4 and 5—but reported on the same scale. This way, educators are assured that a change in scores represents a change in student achievement instead of differences in the tests themselves.

Growth Metrics

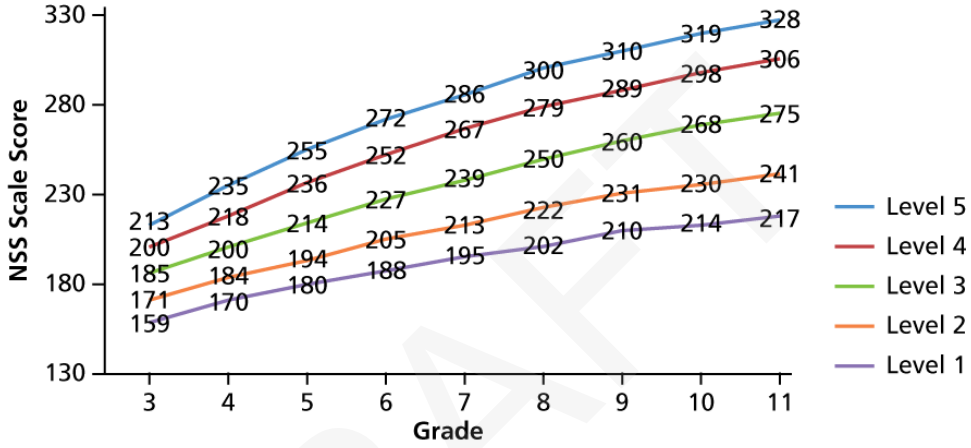
Growth metrics that allow for the longitudinal estimation and tracking of growth are a necessity. The metrics need to be anchored in terms of both content and student performance within and between grades. Three growth metrics are an integral part of the Iowa Growth Model, and all three are expressed in terms of the NSS scale as indicated in Table 4.

Table 4: Growth Metrics Associated with the Iowa Growth Model

Iowa Growth Metric	Notation	Related Terminology
Expected Growth	$NSS_2 \mid NSS_1$	Estimated Growth
Observed Growth	$NSS_2 - NSS_1$	Gain Score Change
Observed - Expected	$NSS_2 - (NSS_2 \mid NSS_1)$	Value-Added

Expected Growth. The relationship between national standard score (NSS) and national percentile rank (NPR) was illustrated previously in Figure 5. Relationships like the one illustrated define, for any student at any level of achievement in one grade, the expected NSS in a subsequent grade (Cunningham, Welch, Dunbar, 2013). When a student has grown as much as expected since the previous year, this student is keeping pace with other students in the nation who started at the same achievement level. The growth chart in Figure 6 consists of a series of curves that illustrate the typical pace of performance for five different students who started in 3rd grade at different achievement levels. For each of these students, the expected NSS for subsequent years is identified. In Table 4, the notation $NSS_2 | NSS_1$ (meaning NSS at time 2 given NSS at time 1) is used to represent the expected score.

Figure 6: Expected Growth Curves for Five Observed Scores (Levels 1 to 5)



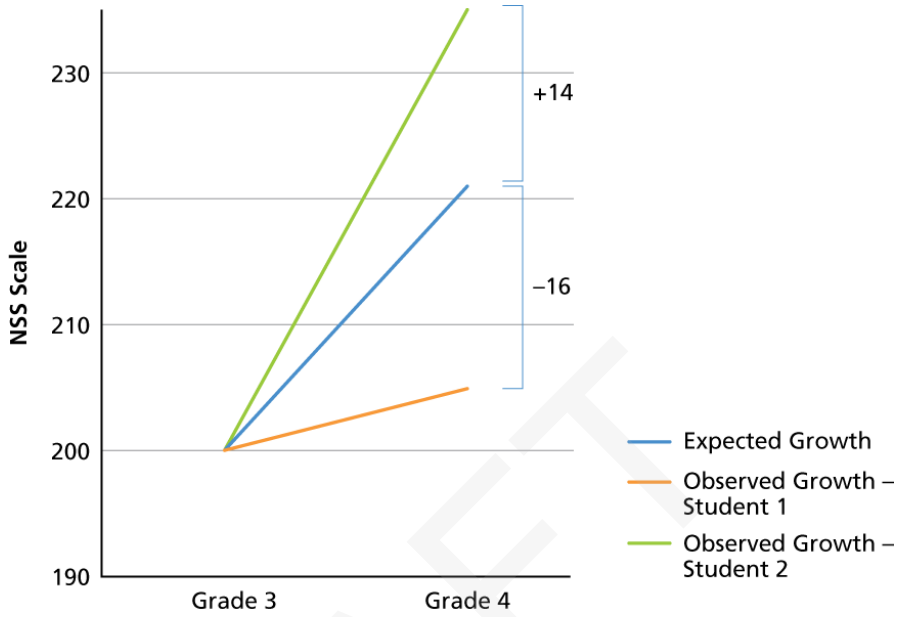
Observed Growth. The observed growth is simply the difference between the second NSS and the first NSS. Observed growth reflects the change in a student's performance between two points in time on the NSS scale. Observed growth is the absolute change in student performance between two time points. These two time points can be from one year to the next, from fall to spring in the same school year, or across multiple years. The sign and magnitude of observed growth are important in indicating a student's change in performance (Castellano and Ho, 2013, p. 36). The sign indicates if the gain is positive, signifying improvement, or negative, signifying decline whereas the magnitude indicates how much the student has changed.

Observed –Expected. The difference between the observed NSS and the expected NSS (given a student's starting point) is frequently described as a "value-added" score. It is the increment of growth that is different from what was expected. As with observed growth, the sign of this quantity is important. If it is positive, then the student has exceeded expectations in growth. If it is zero, then the student has met the expectations in growth. When the quantity is negative, then the student has failed to meet the expectations for growth.

Figure 7 illustrates the relationship among these metrics. Two students were assessed in the fall of 3rd grade, and the *observed* reading score for both students was 200. For these two students and all other students with an NSS of 200 in the fall of 3rd grade, the Iowa Growth Model says that their *expected* NSS in fall of grade 4 is 221. One of the two students obtained

an NSS of 205 in 4th grade, 16 points short of the *expected* NSS of 221 and failed to meet the growth expectation. The other student obtained an NSS of 235 in 4th grade, a 14 point gain over the *expected* NSS of 221. This student exceeded the growth expectation.

Figure 7: Growth Example for Two Students between Grades 3 and 4



Data Requirements and Properties of Measures

The Iowa Growth Model supports multiple approaches to the measurement and evaluation of growth. The fundamental data requirement is a test score on the same scale at two points in time. The NSS is a meaningful metric because it is designed to place students on a developmental continuum in the domain of interest and the scale spans the continuum of learning. In addition the typical magnitude of growth from one grade to the next provides a frame of reference for comparisons of the amount of growth observed in groups of students.

Relationship to other Growth Models

The term "growth model" is used in many achievement contexts, and its meaning is often ambiguous. Ostensibly different growth models may support similar or very different interpretations depending on the statistical foundation of the model and the metrics used to report its results. The results of the Iowa Growth Model have been compared to two "conditional growth" models using two large (state-level) cohorts of students between grade 5 and grade 6 and again between grade 6 and grade 7.

The first conditional growth model is based on the Student Growth Percentile (SGP) metric that describes the rank-order of students in growth relative to peers with similar past test scores. (Betebenner, 2009). The SGP metric relies on quantile regression, conditioning on prior achievement to describe the rank order of the current achievement of students. The second conditional growth model is based on the Percentile Rank of Residuals (PRR) metric which is a ranking of simple differences between observed scores and scores predicted from a linear regression of the current test score on the past score in the same subject area (Castellano, 2012).

Table 5 summarizes the means, standard deviations and sample sizes for the students in grades 5, 6, and 7 in the student cohorts used in the analysis. The mean NSSs in these cohorts represent average achievement in the neighborhood of the 55th to 60th percentile nationally, and the SDs are representative of the variability in the national probability sample of the 2010–2011 norming of the *Iowa Assessments*.

Table 5: Means, Standard Deviations (SD) and Sample Sizes (N) for Growth Metric Comparative Study

Grade	Mathematics			Reading		
	Mean NSS	SD	N	Mean NSS	SD	N
Grade 5	222	24.6	23,452	225	28.6	23,511
Grade 6	232	28.3	27,024	231	32.0	27,046
Grade 7	250	30.6	24,024	245	34.2	27,046

The correlations across grades for the Mathematics and Reading assessments are provided in Table 6. These values are typical of correlations in matched cohorts on assessments that measure a well-defined general achievement construct. They are in the neighborhood of values obtained for test-retest reliability and provide strong support for the quantile and linear regressions need to obtain SGPs and PRRs as indicators of growth.

Table 6: Correlations between Years in Mathematics and Reading

Grade	Mathematics			Reading		
	Grade 5	Grade 6	Grade 7	Grade 5	Grade 6	Grade 7
Grade 5	1.00	–	–	1.00	–	–
Grade 6	.84	1.00	–	.79	1.00	–
Grade 7	.81	.85	1.00	.77	.80	1.00

Comparisons between the results from the Iowa Growth Model and the SGP and PRR approaches are provided in Table 4.7 in terms of correlations between growth indicators. These correlations describe the consistency with which the Iowa Growth Model ranks student growth as compared to the SGP and PRR metrics. In both Mathematics and Reading, these results show that the Iowa Growth Model produces measures of student growth that are virtually identical to those of the other growth metrics.

Table 7: Correlations between Iowa Growth Model, SGP and PRR Metrics

	Iowa Growth Model	
	Mathematics	Reading
Student Growth Percentile (1 prior year)	.98	.97
Percentile Rank of Residuals	.99	.97

Concurrent Validity

Concurrent validity coefficients are presented in the form of correlations between scores on the *Iowa Assessments* Form E and (1) scores on *Cognitive Abilities Test* Form 7 (*CogAT*) and (2) scores on the *Iowa Tests of Basic Skills* and *Iowa Tests of Educational Development* (*ITBS* and *ITED*) Form A.

Form E/*CogAT* Correlations

It is clear from these tables that the highest correlation (with the exception of the Mathematics tests) is provided by the *CogAT* Composite score or the score from the Verbal Battery. The lowest correlations, indicating the least overlap between achievement and the cognitive skills measured, tend to involve the skills tests in the *Iowa Assessments* (for example, Computation and certain tests in the primary levels and grades) and the *CogAT* Form 7 Non-Verbal Battery. One interpretation of the lower correlations in Table 8 is that they represent evidence of discriminant validity.

Average correlations with the *Iowa Assessments* Levels 5/6–17 Complete Composite and *CogAT* Form 7 are .77 for the Verbal Battery, .71 for the Quantitative Battery, .64 for the Nonverbal Battery, and .80 for the *CogAT* Form 7 Composite. Clearly, the relationship is substantial in all cases; however, the correlations are not so high as to suggest that the achievement and ability measures lack discriminant validity.

Table 8: *Iowa Assessments* Form E and *CogAT* Form 7 Correlations

Level (Grade)	N	CogAT	R	V	CW	WE	ET	WA	Li	M	CP	MT	CT	SS	SC	CC
5/6 (1)	1527	Verbal	.42	.48	–	.53	.60	.47	.53	.58	–	–	–	–	–	.64
		Quantitative	.41	.36	–	.46	.52	.42	.44	.56	–	–	–	–	–	.59
		Non-Verbal	.40	.35	–	.46	.51	.42	.45	.55	–	–	–	–	–	.58
		Composite	.47	.44	–	.54	.61	.50	.53	.64	–	–	–	–	–	.68
7 (1)	1557	Verbal	.47	.43	–	.59	.63	.47	.52	.59	–	–	–	–	–	–
		Quantitative	.44	.30	–	.50	.54	.42	.43	.56	–	–	–	–	–	–
		Non-Verbal	.43	.32	–	.50	.53	.44	.42	.54	–	–	–	–	–	–
		Composite	.51	.40	–	.61	.64	.50	.52	.63	–	–	–	–	–	–
8 (2)	3057	Verbal	.56	.54	–	.58	.62	.54	.54	.63	.48	.64	.68	.54	.50	.69
		Quantitative	.49	.48	–	.53	.55	.51	.44	.60	.52	.62	.63	.38	.39	.60
		Non-Verbal	.52	.48	–	.54	.57	.54	.48	.59	.50	.61	.64	.45	.46	.63
		Composite	.58	.56	–	.61	.65	.60	.55	.68	.55	.70	.72	.52	.50	.72
9 (3)	2096	Verbal	.73	.73	.66	.69	.80	.65	.52	.68	.42	.67	.80	.70	.67	.81
		Quantitative	.56	.56	.60	.57	.66	.54	.42	.67	.51	.70	.73	.56	.54	.72
		Non-Verbal	.55	.53	.51	.53	.60	.51	.43	.63	.45	.64	.66	.56	.53	.66
		Composite	.69	.68	.66	.67	.77	.63	.51	.74	.51	.75	.81	.68	.65	.81
10 (4)	2814	Verbal	.77	.76	.69	.71	.81	–	–	.73	.51	.72	.82	.74	.73	.83
		Quantitative	.64	.61	.63	.62	.69	–	–	.75	.59	.77	.76	.64	.65	.75
		Non-Verbal	.59	.56	.56	.58	.63	–	–	.68	.48	.67	.68	.60	.61	.69
		Composite	.75	.72	.70	.71	.79	–	–	.81	.58	.80	.84	.74	.74	.84
11 (5)	2826	Verbal	.78	.77	.71	.74	.82	–	–	.73	.50	.72	.82	.75	.76	.84
		Quantitative	.63	.60	.65	.64	.70	–	–	.76	.61	.78	.77	.62	.62	.75
		Non-Verbal	.58	.53	.56	.58	.63	–	–	.68	.49	.68	.69	.58	.60	.69
		Composite	.75	.72	.74	.74	.80	–	–	.81	.59	.81	.85	.73	.74	.85
12 (6)	2444	Verbal	.78	.79	.67	.74	.82	–	–	.71	.49	.69	.80	.73	.73	.82
		Quantitative	.60	.59	.63	.63	.68	–	–	.77	.64	.78	.76	.58	.62	.74
		Non-Verbal	.53	.52	.54	.56	.60	–	–	.64	.46	.63	.64	.48	.52	.62
		Composite	.73	.73	.70	.73	.80	–	–	.79	.56	.78	.83	.66	.68	.82
13 (7)	1864	Verbal	.76	.74	.64	.69	.78	–	–	.66	.41	.63	.77	.67	.67	.78
		Quantitative	.61	.52	.62	.61	.67	–	–	.77	.59	.78	.76	.56	.59	.74
		Non-Verbal	.51	.46	.52	.52	.57	–	–	.64	.46	.63	.64	.48	.52	.62
		Composite	.73	.67	.68	.70	.78	–	–	.79	.56	.78	.83	.66	.68	.82

Continued on next page…

Table 8 (*continued*): *Iowa Assessments* Form E and *CogAT* Form 7 Correlations

Level (Grade)	N	CogAT	R	V	CW	WE	ET	WA	Li	M	CP	MT	CT	SS	SC	CC
14 (8)	1895	Verbal	.76	.77	.66	.70	.79	–	–	.69	.50	.68	.78	.71	.70	.80
		Quantitative	.62	.56	.64	.63	.68	–	–	.79	.63	.79	.78	.57	.61	.75
		Non-Verbal	.58	.52	.55	.55	.61	–	–	.68	.53	.68	.68	.53	.59	.67
		Composite	.75	.71	.70	.72	.79	–	–	.82	.63	.82	.85	.69	.72	.85
15 (9)	1940	Verbal	.72	.75	–	.66	.78	–	–	.63	.51	.66	.77	.66	.66	.79
		Quantitative	.56	.53	–	.56	.62	–	–	.67	.65	.74	.73	.53	.56	.71
		Non-Verbal	.47	.46	–	.51	.55	–	–	.58	.52	.62	.63	.47	.51	.62
		Composite	.68	.69	–	.68	.76	–	–	.74	.64	.78	.82	.65	.68	.82
16 (10)	2002	Verbal	.71	.68	–	.68	.76	–	–	.63	.47	.63	.74	.63	.63	.75
		Quantitative	.58	.5	–	.60	.67	–	–	.69	.63	.74	.75	.56	.58	.74
		Non-Verbal	.54	.50	–	.55	.59	–	–	.62	.53	.65	.67	.52	.56	.66
		Composite	.70	.66	–	.70	.77	–	–	.73	.62	.77	.82	.65	.67	.82
17 (11)	2188	Verbal	.65	.69	–	.64	.70	–	–	.59	.47	.60	.70	.64	.64	.72
		Quantitative	.55	.51	–	.58	.60	–	–	.65	.62	.70	.70	57	.57	.68
		Non-Verbal	.49	.47	–	.50	.53	–	–	.57	.49	.58	.60	.48	.52	.59
		Composite	.65	.64	–	.66	.70	–	–	.69	.61	.72	.77	.65	.67	.77

Iowa Assessments Form E and *ITBS/ITED* Form A Correlations

As part of the National Comparison Study, some students were administered both Form E of the *Iowa Assessments* and Form A of the *ITBS or ITED.* These data were used to link NSS's on the two forms and to examine the strength of the relationship between forms. Studies of internal structure discussed earlier and detailed in Chen, Welch and Dunbar (2013) suggest a certain degree of comparability in underlying achievement constructs. The concurrent validity coefficients from the matched Form E and Form A data are reported in Table 9. In English language arts and mathematics, the coefficients are generally in the .75 to .85 range except at grades 1 and 2, where they tend to be slightly lower. Students taking alternate forms of the *Iowa Assessments* are rank-ordered in a highly similar fashion, suggesting that when administrative conditions are monitored appropriately, the *Iowa Assessments* produce scores that are dependable in the sense that they are minimally affected by factors beyond the control of test administrators.

Table 9: *Iowa Assessments*, Form E and *ITBS/ITED* Form A Correlations

Test Level (Grade)	N	R	L	M	SS	SC	V	SP	CP	PC	MC	WA	Li
Level 7 (Grade 1)	1738	.86	.77	.79	.63	.60	.83	–	–	–	.70	.72	.68
Level 8 (Grade 2)	1068	.81	.82	.84	.61	.66	.78	–	–	–	.70	.75	.66
Level 9 (Grade 3)	965	.84	.79	.83	.79	.74	.83	.78	.75	.72	.77	.72	.58
Level 10 (Grade 4)	2072	.82	.79	.84	.76	.76	.84	.83	.76	.77	.75	–	–
Level 11 (Grade 5)	2084	.82	.81	.84	.79	.78	.83	.82	.79	.80	.76	–	–
Level 12 (Grade 6)	1163	.83	.81	.85	.80	.75	.84	.84	.77	.78	.75	–	–
Level 13 (Grade 7)	1041	.80	.84	.86	.79	.78	.83	.84	.81	.81	.73	–	–

Continued on next page…

Table 9 (*continued*): *Iowa Assessments*, Form E and *ITBS/ITED* Form A Correlations

Test Level (Grade)	N	R	L	M	SS	SC	V	SP	CP	PC	MC	WA	Li
Level 14 (Grade 8)	1184	.84	.84	.89	.80	.74	.86	.86	.81	.84	.76	–	–
Level 15 (Grade 9)	784	.72	.77	.75	.76	.70	.78	–	–	–	.61	–	–
Level 16 (Grade 10)	583	.76	.84	.78	.75	.78	.85	–	–	–	.61	–	–
Level 17 (Grade 11)	704	.58	.61	.51	.62	.55	.77	–	–	–	.62	–	–

Note: Tests with blank cells are not given in the level in which the blank cells appear.

Other Validity Considerations

Universal Design

The principles of universal design for assessments provide guidelines for the test development process intended to ensure that no test takers are unduly disadvantaged owing to a special need, incomplete language mastery, or membership in any demographic group. Universal design in the development of assessment materials involves aspects of presentation in both paper-based and computer-based modes of administration to enhance accessibility and clarity for all examinees Universal design principles are not intended to make any test easier for a given subgroup, but only to remove the effects of construct-irrelevant variance on test scores. Ease of navigation of test materials; clarity of typeface, graphics, and page layout; and respect for the diversity of the test-taking population in the nature of the materials presented are some examples of universal design principles for assessments.

An independent, comprehensive universal design review of page layouts, color schemes, and other factors in the design and presentation of materials for the *Iowa Assessments* was conducted by the National Center of Educational Outcomes (NCEO) at the University of Minnesota. A review panel consisting of experts in fields such as special education, English language learning, assessment of students with special needs, and education in urban areas produced a report that helped guide final decisions in the publication of the *Iowa Assessments*. This review was conducted prior to the National Comparison Study and development of norms and score conversions from the 2010–11 national probability sample.

Color Blindness

Informational graphics for the final publication of the *Iowa Assessments* were subject to a through a composition check to ensure coherency and effective color contrast for students with a color vision deficiency. Art was processed through a color blindness simulator that emulates red-blind, green-blind, and blue-blind conditions (protanopia, deuteranopia and tritanopia, respectively). If required, color was adjusted and then resubmitted to the simulator for validation.

Graphics were validated as acceptable for color-blind students using Vischeck (http://www.vischeck.com). Vischeck is an online or downloadable color blindness simulator that renders images as they would appear to individuals with protanopia, deuteranopia, or tritanopia. Using these simulations as a guide, any art requiring modification was revised by

choosing patterns and/or color contrast that were acceptable for individuals with a color vision deficiency. All revised art and graphics were re-tested using Vischeck to ensure color contrast was sufficient for the simulated conditions.

Text Complexity and Readability

The best way to determine the difficulty of a large-scale assessment is to examine item and test data that indicate the average levels of performance obtained by the examinees for whom the assessment is intended. The difficulty data for items, skill domains, and tests in the *Iowa Assessments* are reported in *Content Classifications Guide* for Levels 5/6–14 and Levels 15–17/18. Of the various factors that influence difficulty, text complexity, sometimes called readability, is the focus of much attention.

The readability of written materials is measured in several ways. An expert may judge the grade level of a reading passage based on perception of its complexity. The most common method of quantifying these judgments is to use one of an ever-expanding array of text complexity or readability algorithms (see Nelson, Perfetti, Liben & Liben, 2012, for a recent review and comparison of text complexity measures). These measures use word frequency, word and sentence length, and other features of text (for example, unusual letter patterns, subordination, sentence cohesion, and so forth) and usually produce a single measure of text complexity or readability for each block of text analyzed. Nelson, et al (2012) found "impressively high" (p. 3) correlations between all the measures they studied and student performance on the standardized tests from which they drew text for the analysis. This finding suggests that test assembly practices that use item difficulty data from field testing to gauge the appropriateness of assessment materials for a given grade level simultaneously monitor text complexity such that it is appropriate for the range of reading levels in the student population.

The virtue of text complexity formulas is objectivity. Their shortcoming is failure to account for qualitative factors that influence how easily a reader comprehends written material. Such factors may include the organization and cohesiveness—some approaches include these elements—of a selection, cognitive complexity of the concepts presented, amount of knowledge a reader is expected to bring to the selection, clarity of new information, or interest level of the material to its audience. These other factors are likely to influence student performance on assessments, so empirical measures of item and test difficulty remain an important aspect of any evaluation of text complexity or readability.

Review of Materials for the *Iowa Assessments*. Consistent with the recommendations of the Common Core State Standards, three different dimensions are used to describe the text complexity of the *Iowa Assessments* in the areas of reading, language arts, social studies and science. These dimensions are qualitative, quantitative, and reader/task oriented in nature, and Table 10 on the next page summarizes the type of information available to help evaluate each dimension. All three dimensions are equally important in the assembly of operational forms; they are used to provide a range of text complexity within a form and across forms to help ensure that the forms are as comparable as possible.

All text-based materials are reviewed by testing and content experts for the four different aspects of the qualitative dimension including level of meaning or purpose, structure, language conventionality and clarity. Each test form is being assembled to include a balance

of the range of these dimensions. For example, Form E was assembled to include a range of text types of increasing complexity and sophistication at test level increased. The quantitative dimensions are evaluated through a combination of text-based indices (for example, Lexiles and traditional readability indices) and national passage-based statistics that addresses the relative difficulty of these materials for examinees in the intended grade. In addition, all passages were reviewed as they were developed and selected for accessibility, appropriateness of test complexity, and interest level.

Table 10: Text Complexity Considerations

Dimension	Considerations for *Iowa Assessments*				
Qualitative Dimension	**Test**	**Reading**	**Written Expression**	**Social Studies**	**Science**
	Levels of Meaning or Purpose	Includes a variety of literary and informational texts from simple meaning to multiple meanings	Includes a variety of literacy and informational texts from explicitly stated to implicitly stated		
	Structure	Includes a variety of texts from simple to highly complex	Includes a variety of texts from simple to highly complex	Graphics and figures range from simple to complex	Graphics and figures range from simple to complex
	Language Conventionality and Clarity	Texts rely on a range of language conventionality and clarity from literal to figurative. Texts are balanced to represent this range within each assembled form.			
	Knowledge Demands	No assumptions about readers' life experiences	No assumptions about readers' life experience	Background content knowledge assumed	Background content knowledge assumed
Quantitative Dimension	• Lexile scores for all text-based stimuli aligned to grade-level ranges established by MetaMetrics • Traditional readability indices for all text-based stimuli based on word length, frequency and complexity • Item-level and form-level difficulty indices collected from a nationally representative sample of students in grades K-12.				
Reader and Task Considerations	• Student difficulty levels collected on nationally representative samples of students in relevant grades. • Professional judgments from educators on the appropriateness of the passages and stimuli included in the assembled forms				

Reliability

In Brief

This part of the guide reports several different estimates of reliability that can help users make informed judgments about the consistency of *Iowa Assessments* scores. Data presented in this part of the guide address the means, standard deviations (SD), and standard errors of measurement (SEM) for raw scores (RS) and National Standard Scores (NSS). Several approaches to the assessment of reliability and sources of variance in observed scores are also presented as well as standard errors of measurement for selected scores levels, also known as conditional SEMs.

Methods of Determining, Reporting, and Using Reliability Data

A soundly planned, carefully constructed and comprehensive large-scale assessment represents the most accurate and dependable measure of student achievement available to parents, teachers, and school officials. Many subtle, extraneous factors that contribute to unreliability and bias in human judgments have little or no effect on scores from carefully developed assessments. In addition, other factors that contribute to apparent inconsistency in student performance can be effectively minimized in the assessment situation: temporary changes in student motivation, health, and attentiveness; minor distractions inside and outside the classroom; limitations in number, scope, and comparability of the available samples of student work; and misunderstanding by students of what the teacher expects of them. The greater effectiveness of a well-constructed achievement test in controlling these factors—compared to informal evaluations of the same achievement—is evidenced by the higher reliability of the test.

Test reliability can be quantified by a variety of statistical data, but such data reduce to two basic types of indices. The first of these indices is the reliability coefficient. In numerical value, the reliability coefficient is between .00 and .99; for standardized assessments it is generally between .60 and .95. The closer the coefficient approaches the upper limit, the greater the freedom of the scores from the influence of factors that temporarily affect student performance and obscure real differences in achievement. This ready frame of reference for reliability coefficients is deceptive in its simplicity, however. It is impossible to conclude whether a value such as .75 represents a "high" or "low" or "satisfactory" or "unsatisfactory" reliability. Only after a coefficient has been compared to those of *equally valid* and *equally practical* alternative assessments can such a judgment be made. In practice, there is always a degree of uncertainty regarding the terms "equally valid" and "equally practical," so the reliability coefficient is rarely free of ambiguity. Nonetheless, comparisons of reliability coefficients for alternative approaches to assessment can be useful in determining the relative stability of the resulting scores.

The second of the statistical indices used to describe test reliability is the standard error of measurement. This index represents a measure of the net effect of all factors leading to inconsistency in student performance and to inconsistency in the interpretation of that performance. The SEM can be understood by a hypothetical example. Suppose a group of students at the same achievement level in reading were to take the same reading test on two occasions. Despite their equal reading ability, they would not all get the same score both times. Instead, their scores would range across an interval. A very few would get much higher scores than expected given their achievement level and a few much lower; the majority would get scores quite close to their actual achievement level. Such variation in scores would be attributable to differences in motivation, attentiveness, and other situational factors. The SEM is an index of the typical range or variability of the scores observed for students regardless of their level of achievement. It tells the degree of precision in placing a student at a point on the score scale used for reporting assessment results.

There is, of course, no way to know just how much a given student's achievement may have been under- or over-estimated from a single administration of a test. We may, however, make reasonable estimates of the amount by which the achievement of students in a particular reference group has been mismeasured. For about two-thirds of the examinees, the scores obtained are "correct" or accurate to within one SEM of the observed score. For 95 percent of the students, the scores are accurate to within two standard errors, and for more than 99 percent, the scores are accurate to within three standard error values.

Two methods of estimating reliability were used to obtain the summary statistics provided in the following two sections of this guide. The first method employed internal-consistency estimates using Kuder-Richardson Formula 20 (K-R20). Reliability coefficients derived by this technique were based on data from the entire national standardization sample and are reported for both fall and spring administrations. The coefficients for Form E of the *Iowa Assessments* Complete Battery are reported here. Coefficients for Form F and for Forms E and F of the *Iowa Assessments* Survey Battery are available from the publisher.

The second method provided estimates of reliability based on two testing occasions. Alternate-forms reliability for Form E of the *Iowa Assessments* and Form A of the *ITBS*/*ITED* were estimated from the fall 2010 equating of those forms. In addition, test-retest reliability was estimated with data from the 2011–2012 comparability study of Form E in paper-based and computer-based modes of administration.

The SEM measures the net effect of all factors leading to inconsistency in student test scores and to inconsistency in score interpretation. It is reported as the typical amount by which a student's observed score may range from one testing occasion to another. The conditional SEM (CSEM) gives similar information, but rather than gauging the typical range, it provides a range that is tailored to a specific level of achievement (Feldt & Brennan, 1989; Haertel, 2006).

The reliability data presented on the following pages are based on Kuder-Richardson Formula 20 (K-R 20). The means, standard deviations, and standard errors of measurement are shown in Table 11 in the raw score metric and the National Standard Score metric for both fall and spring administrations of the *Iowa Assessments*.

Table 11: Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 5/6		Reading	Language	Vocabulary	**ELA Total**	Word Analysis	Listening	**Extended ELA**	Mathematics	**Complete Composite**	**Extended Complete Composite**
		R	**L**	**V**	**ET**	**WA**	**Li**	**XET**	**M**	**CC**	**XCC**
Number of Items		**34**	**31**	**27**		**33**	**27**		**35**		
Fall–Grade 1											
	Mean	17.8	20.1	17.6	–	26.9	17.0	–	22.6	–	–
RS	SD	8.1	4.8	3.6	–	4.9	4.3	–	6.2	–	–
	SEM	2.5	2.3	2.2	–	2.0	2.3	–	2.4	–	–
	Mean	139.1	137.3	138.1	138.0	138.9	138.1	138.2	138.3	138.2	138.3
SS	SD	10.2	9.6	16.0	9.9	15.9	11.9	12.3	11.3	10.9	10.9
	SEM	3.2	4.6	9.7	3.0	6.5	6.3	2.5	4.5	2.7	2.6
K-R 20		.903	.770	.625	.907	.836	.724	.958	.844	.939	.945
Spring–Grade K											
	Mean	11.6	17.4	16.0	–	24.5	14.4	–	18.4	–	–
RS	SD	5.8	5.8	3.6	–	5.7	4.3	–	5.8	–	–
	SEM	2.5	3.0	2.3	–	2.2	2.4	–	2.6	–	–
	Mean	131.3	130.4	131.1	130.8	131.5	130.8	130.9	130.7	130.8	130.8
SS	SD	7.4	8.5	15.0	8.5	14.3	10.8	11.1	9.8	9.8	9.8
	SEM	3.6	4.4	9.7	2.9	5.5	6.0	2.4	4.3	2.6	2.5
K-R 20		.810	.740	.580	.882	.853	.690	.954	.804	.929	.936

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 7		Reading	Language	Vocabulary	**ELA Total**	Word Analysis	Listening	**Extended ELA**	**Mathematics**			**Core Composite**	Extended English Language Arts Total	Core Composite with ET and M	Core Composite with XET and M	Science	Social Studies	**Complete Composite**	Complete Composite with XET and MT	Complete Composite with ET and M	Complete Composite with XET and M
									Mathematics	Computation	**Math Total**										
		R	**L**	**V**	**ET**	**WA**	**Li**	**XET**	**M**	**MC**	**MT**	**CT**	**XCT**	**CT-**	**XCT-**	**SC**	**SS**	**CC**	**XCC**	**CC-**	**XCC-**
Number of Items		**35**	**34**	**26**		**32**	**27**		**41**	**25**						**29**	**29**				
Fall–Grade 2																					
	Mean	26.8	23.6	17.9	–	25.9	20.3	–	29.1	18.8	–	–	–	–	–	22.9	23.0	–	–	–	–
RS	SD	6.9	6.7	5.9	–	4.6	4.1	–	6.0	4.7	–	–	–	–	–	3.2	3.4	–	–	–	–
	SEM	2.1	2.3	1.9	–	2.0	2.0	–	2.4	1.9	–	–	–	–	–	1.9	1.9	–	–	–	–
	Mean	158.9	158.1	157.5	158.3	159.2	156.9	158.2	157.0	154.2	156.1	157.2	157.1	157.6	157.6	157.4	157.8	157.3	157.3	157.6	157.6
SS	SD	16.3	15.1	19.0	15.1	20.4	14.8	14.0	14.8	9.9	12.8	13.4	13.4	13.8	13.8	18.3	16.3	13.0	13.0	13.2	13.2
	SEM	5.0	5.3	6.22	3.3	8.9	7.1	2.9	6.0	4.0	3.8	2.5	2.4	3.1	3.0	11.0	9.3	2.9	2.9	3.2	3.2
K-R 20		.906	.879	.893	.953	.810	.768	.957	.836	.835	.911	.965	.968	.949	.952	.641	.676	.948	.950	.941	.943
Spring–Grade 1																					
	Mean	23.8	20.0	15.7	–	24.3	18.6	–	26.1	16.8	–	–	–	–	–	21.5	21.5	–	–	–	–
RS	SD	7.6	5.9	6.2	–	4.8	4.3	–	5.7	5.0	–	–	–	–	–	3.4	3.7	–	–	–	–
	SEM	2.3	2.6	2.1	–	2.2	2.2	–	2.7	2.1	–	–	–	–	–	2.1	2.1	–	–	–	–
	Mean	152.2	149.9	150.9	150.8	152.2	150.4	151.0	150.3	150.1	150.2	150.5	150.6	150.6	150.6	149.8	151.1	150.5	150.6	150.5	150.6
SS	SD	14.3	11.3	18.0	11.3	18.4	13.5	12.6	13.6	9.3	11.2	12.2	12.2	12.7	12.7	16.8	15.2	11.8	11.8	12.3	12.3
	SEM	4.4	5.0	6.1	3.1	8.4	6.7	2.7	4.9	3.9	3.5	2.3	2.2	2.9	2.8	10.2	8.6	2.8	2.7	3.0	2.9
K-R 20		.904	.808	.884	.927	.794	.750	.953	.867	.829	.901	.963	.966	.948	.951	.630	.679	.946	.947	.941	.942

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 8		Reading	Language	Vocabulary	**ELA Total**	Word Analysis	Listening	**Extended ELA**	**Mathematics**			**Core Composite**	Extended English Language Arts Total	Core Composite with ET and M	Core Composite with XET and M	Science	Social Studies	**Complete Composite**	Complete Composite with XET and MT	Complete Composite with ET and M	Complete Composite with XET and M
									Mathematics	Computation	**Math Total**										
		R	**L**	**V**	**ET**	**WA**	**Li**	**XET**	**M**	**MC**	**MT**	**CT**	**XCT**	**CT-**	**XCT-**	**SC**	**SS**	**CC**	**XCC**	**CC-**	**XCC-**
Number of Items		**38**	**42**	**26**		**33**	**27**		**46**	**27**						**29**	**29**				
Fall–Grade 3																					
	Mean	29.1	32.0	17.8	–	26.6	20.1	–	34.5	21.2	–	–	–	–	–	21.3	22.1	–	–	–	–
RS	SD	6.8	7.5	4.6	–	4.8	4.1	–	7.0	3.7	–	–	–	–	–	3.9	4.1	–	–	–	–
	SEM	2.3	2.4	2.0	–	2.0	2.0	–	2.5	2.0	–	–	–	–	–	2.1	2.0	–	–	–	–
	Mean	177.5	177.0	175.4	176.9	177.6	174.6	176.6	175.3	172.3	174.3	175.6	175.5	176.1	176.0	177.0	176.6	176.0	175.9	176.3	176.2
SS	SD	21.4	19.5	20.6	20.1	25.4	17.3	17.1	18.4	13.9	16.0	16.7	16.7	17.4	17.4	22.5	19.4	17.1	17.1	17.1	17.1
	SEM	7.1	6.4	9.0	4.2	10.4	8.5	3.6	6.0	5.3	4.4	3.1	2.8	3.7	3.5	12.0	9.6	3.3	3.2	3.6	3.5
K-R 20		.890	.893	.808	.955	.833	.761	.956	.892	.857	.925	.967	.971	.955	.959	.714	.757	.963	.965	.957	.958
Spring–Grade 2																					
	Mean	27.0	29.2	16.3	–	25.4	19.2	–	32.0	20.3	–	–	–	–	–	20.0	20.6	–	–	–	–
RS	SD	7.1	7.9	4.6	–	5.1	4.3	–	7.0	3.8	–	–	–	–	–	4.0	4.1	–	–	–	–
	SEM	2.4	2.7	2.1	–	2.1	2.2	–	2.7	2.1	–	–	–	–	–	2.2	2.2	–	–	–	–
	Mean	170.7	169.8	168.6	169.9	171.0	168.2	169.8	168.6	168.3	168.5	169.2	169.1	169.2	169.2	169.7	169.5	169.3	169.3	169.4	169.3
SS	SD	19.6	17.2	19.8	17.2	23.7	16.3	16.1	16.9	13.1	14.7	15.3	15.3	15.9	15.9	21.2	17.8	15.0	15.0	15.2	15.2
	SEM	6.7	5.8	9.1	4.0	9.8	8.3	3.4	5.9	5.4	4.3	2.9	2.7	3.5	3.4	11.5	9.3	3.2	3.1	3.4	3.4
K-R 20		.883	.886	.791	.947	.828	.745	.955	.879	.834	.914	.964	.968	.950	.954	.708	.727	.955	.957	.949	.951

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 9 Grade 3		Reading	Written Expression	Conventions of Writing				Vocabulary	ELA Total	Word Analysis	Listening	Extended ELA
				Spelling	Capitalization	Punctuation	Conventions of Writing Total					
		R	WE	SP	CP	PC	CW	V	ET	WA	Li	XET
Number of Items		41	35	24	20	20		29		33	28	
Fall												
	Mean	23.3	19.1	12.8	9.9	8.8	–	16.2	–	21.0	16.2	–
RS	SD	8.6	7.7	5.0	4.7	4.1	–	6.9	–	5.2	3.7	–
	SEM	2.7	2.6	2.1	1.9	2.0	–	2.3	–	2.5	2.3	–
	Mean	177.5	176.8	175.4	175.1	177.5	174.2	175.4	176.3	177.6	174.6	176.2
SS	SD	21.4	23.9	17.9	23.2	23.6	19.5	20.6	20.1	25.4	17.3	17.1
	SEM	6.8	7.8	7.7	9.4	11.2	5.4	6.9	3.8	12.2	10.8	3.7
K-R 20		.900	.887	.816	.836	.775	.922	.888	.965	.771	.613	.953
Spring												
	Mean	27.0	22.4	15.2	11.8	10.5	–	19.3	–	22.8	18.1	–
RS	SD	8.5	8.0	5.0	5.1	4.4	–	6.8	–	5.4	3.8	–
	SEM	2.6	2.4	2.0	1.8	1.9	–	2.2	–	2.4	2.2	–
	Mean	187.8	188.7	185.8	187.2	188.3	185.2	185.0	187.2	187.2	184.2	186.7
SS	SD	24.5	28.2	20.4	29.2	27.4	22.7	21.6	21.7	28.6	19.2	19.0
	SEM	7.5	8.6	8.1	10.3	11.8	5.8	6.8	4.1	12.7	10.8	3.9
K-R 20		.906	.907	.840	.875	.813	.934	.900	.965	.804	.683	.958

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 9 Grade 3		Mathematics			Core Composite	Core Composite with XET	Core Composite with ET and M	Core Composite with XET and M	Science	Social Studies	Complete Composite	Complete Composite with XET and MT	Complete Composite with ET and M	Complete Composite with XET and M
		Mathematics	Computation	Math Total										
		M	MC	MT	CT	XCT	CT-	XCT-	SC	SS	CC	XCC	CC-	XCC-
Number of Items		50	25						30	30				
Fall														
	Mean	25.6	12.0	–	–	–	–	–	15.6	18.1	–	–	–	–
RS	SD	8.3	5.5	–	–	–	–	–	6.0	6.1	–	–	–	–
	SEM	3.1	2.2	–	–	–	–	–	2.4	2.4	–	–	–	–
	Mean	175.3	172.3	174.3	175.3	175.3	175.8	175.8	177.0	176.6	175.8	175.8	176.1	176.1
SS	SD	18.4	13.9	16.0	16.7	16.7	17.4	17.4	22.5	19.4	17.1	17.1	17.1	17.1
	SEM	6.8	5.5	4.9	3.1	3.1	3.9	3.9	8.8	7.5	2.8	2.8	3.2	3.2
K-R 20		.861	.846	.906	.966	.966	.950	.950	.846	.850	.972	.973	.964	.964
Spring														
	Mean	30.2	17.0	–	–	–	–	–	18.2	20.8	–	–	–	–
RS	SD	8.8	5.8	–	–	–	–	–	6.0	5.8	–	–	–	–
	SEM	3.0	1.9	–	–	–	–	–	2.3	2.2	–	–	–	–
	Mean	185.9	185.4	185.7	186.4	186.2	186.5	186.3	187.4	186.8	186.7	186.5	186.7	186.6
SS	SD	20.5	16.7	17.7	19.1	19.1	19.9	19.9	25.2	21.7	19.9	19.9	20.0	20.0
	SEM	7.0	5.5	5.0	3.2	3.2	4.0	4.0	9.5	8.3	3.0	3.0	3.4	3.4
K-R 20		.884	.891	.920	.971	.972	.959	.960	.858	.853	.977	.977	.971	.971

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 10 Grade 4		Reading	Written Expression	Conventions of Writing				Vocabulary	ELA Total	Mathematics			Core Composite	Core Composite with ET and M	Science	Social Studies	Complete Composite	Complete Composite with ET and M
				Spelling	Capitalization	Punctuation	Conventions of Writing Total			Mathematics	Computation	Math Total						
		R	WE	SP	CP	PC	CW	V	ET	M	MC	MT	CT	CT-	SC	SS	CC	CC-
Number of Items		42	38	27	22	22		34		55	27				34	34		
Fall																		
	Mean	25.9	22.0	15.2	11.6	10.2	–	19.9	–	30.7	15.4	–	–	–	19.0	19.4	–	–
RS	SD	8.8	8.4	5.8	5.0	4.5	–	7.9	–	9.3	5.5	–	–	–	6.4	6.6	–	–
	SEM	2.7	2.6	2.2	2.0	2.0	–	2.5	–	3.2	2.2	–	–	–	2.5	2.5	–	–
	Mean	193.8	195.1	192.2	194.0	195.4	191.9	191.1	193.5	191.8	188.8	190.8	192.1	192.6	193.8	192.6	192.5	192.8
SS	SD	25.9	30.5	22.2	31.4	30.0	24.9	22.5	22.8	21.8	17.4	18.9	20.4	21.2	26.6	23.3	21.4	21.3
	SEM	8.0	9.5	8.5	12.8	13.4	6.7	7.1	4.4	7.6	7.0	5.6	3.6	4.4	10.5	9.0	3.3	3.7
K-R 20		.904	.903	.855	.834	.800	.927	.900	.962	.878	.840	.913	.970	.957	.844	.852	.976	.969
Spring																		
	Mean	28.3	24.2	17.6	12.8	11.5	–	22.8	–	34.6	18.6	–	–	–	21.2	21.9	–	–
RS	SD	8.7	8.5	5.7	5.3	5.0	–	7.7	–	9.5	5.7	–	–	–	6.5	6.7	–	–
	SEM	2.6	2.5	2.1	2.0	2.0	–	2.4	–	3.1	2.1	–	–	–	2.5	2.4	–	–
	Mean	202.6	204.9	202.5	204.0	204.9	201.8	199.9	202.8	201.6	200.7	201.3	202.0	202.2	203.5	202.6	202.4	202.5
SS	SD	28.7	34.6	25.3	36.2	34.4	28.4	23.4	24.4	24.0	20.5	21.1	22.9	23.7	29.2	26.4	23.9	23.9
	SEM	8.6	10.1	9.2	13.6	13.6	7.0	7.2	4.7	7.9	7.5	5.9	3.8	4.6	11.1	9.5	3.5	3.9
K-R 20		.911	.914	.867	.859	.843	.938	.905	.963	.891	.866	.922	.973	.962	.857	.870	.978	.973

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 11 Grade 5		Reading	Written Expression	Conventions of Writing				Vocabulary	**ELA Total**	Mathematics			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
				Spelling	Capitalization	Punctuation	Conventions of Writing Total			Mathematics	Computation	**Math Total**						
		R	**WE**	**SP**	**CP**	**PC**	**CW**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**43**	**40**	**30**	**24**	**24**		**37**		**60**	**29**				**37**	**37**		
Fall																		
	Mean	26.7	24.7	17.6	12.2	11.3	–	22.3	–	34.6	16.9	–	–	–	21.1	20.4	–	–
RS	SD	9.1	8.9	6.4	5.1	4.8	–	8.2	–	10.1	6.0	–	–	–	6.7	7.6	–	–
	SEM	2.7	2.6	2.3	2.1	2.2	–	2.6	–	3.4	2.3	–	–	–	2.6	2.7	–	–
	Mean	207.0	209.8	207.7	209.0	210.3	206.9	205.1	207.6	206.7	204.2	205.9	206.8	207.2	208.5	207.2	207.1	207.4
SS	SD	29.9	36.4	26.8	37.9	36.6	30.5	24.0	25.4	25.6	21.4	22.3	24.2	25.4	30.6	28.2	25.2	25.1
	SEM	9.0	10.8	9.9	15.6	16.4	8.2	7.6	5.0	8.6	8.0	6.3	4.0	5.0	11.9	9.9	3.7	4.2
K-R 20		.909	.912	.864	.830	.799	.928	.900	.961	.888	.861	.920	.972	.962	.848	.877	.978	.972
Spring																		
	Mean	29.0	26.5	19.5	13.5	12.5	–	25.1	–	38.0	19.6	–	–	–	23.1	22.9	–	–
RS	SD	9.1	9.0	6.3	5.3	5.2	–	8.3	–	10.5	6.2	–	–	–	6.9	7.9	–	–
	SEM	2.6	2.5	2.2	2.1	2.1	–	2.4	–	3.3	2.2	–	–	–	2.5	2.6	–	–
	Mean	215.5	218.9	216.9	218.5	219.3	216.1	214.0	216.5	215.8	215.3	215.6	216.0	216.1	217.9	217.0	216.5	216.6
SS	SD	32.2	40.1	29.3	41.0	40.0	33.3	25.5	27.3	27.9	24.7	24.8	26.3	27.2	33.3	31.2	27.5	27.7
	SEM	9.2	11.3	10.4	15.8	16.1	8.2	7.5	5.2	8.6	8.5	6.4	4.1	5.0	12.2	10.2	3.8	4.3
K-R 20		.918	.921	.874	.851	.838	.939	.913	.964	.904	.881	.933	.975	.966	.866	.894	.981	.976

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 12 Grade 6		Reading	Written Expression	Conventions of Writing				Vocabulary	**ELA Total**	Mathematics			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
				Spelling	Capitalization	Punctuation	Conventions of Writing Total			Mathematics	Computation	**Math Total**						
		R	**WE**	**SP**	**CP**	**PC**	**CW**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**44**	**43**	**32**	**25**	**25**		**39**		**65**	**30**				**39**	**39**		
Fall																		
	Mean	29.2	26.5	18.9	12.5	12.4	–	23.4	–	37.9	17.8	–	–	–	20.6	22.4	–	–
RS	SD	8.9	8.7	6.9	4.9	4.8	–	7.8	–	11.5	6.3	–	–	–	7.2	8.0	–	–
	SEM	2.7	2.8	2.4	2.2	2.2	–	2.7	–	3.5	2.3	–	–	–	2.8	2.8	–	–
	Mean	220.0	223.3	221.5	223.1	224.0	220.6	219.2	221.0	220.5	219.3	220.1	220.6	220.8	221.8	221.5	221.0	221.1
SS	SD	33.4	41.7	30.3	42.1	41.5	34.8	26.3	28.2	28.9	25.7	25.8	27.5	28.2	34.6	32.5	28.5	28.5
	SEM	10.0	13.2	10.5	19.1	19.1	9.4	9.0	5.9	8.9	9.4	6.7	4.5	5.3	13.4	11.2	4.2	4.6
K-R 20		.910	.899	.881	.805	.789	.926	.883	.956	.906	.866	.933	.974	.964	.851	.882	.979	.974
Spring																		
	Mean	30.1	27.9	20.5	13.3	13.3	–	25.5	–	41.0	19.7	–	–	–	22.4	24.1	–	–
RS	SD	8.8	9.0	6.8	5.1	5.2	–	7.9	–	11.7	6.7	–	–	–	7.5	8.4	–	–
	SEM	2.6	2.7	2.3	2.2	2.1	–	2.6	–	3.4	2.2	–	–	–	2.7	2.7	–	–
	Mean	227.3	230.8	229.5	231.1	232.4	228.7	226.7	228.6	228.7	228.4	228.6	228.6	228.7	230.7	229.6	229.1	229.2
SS	SD	35.3	45.0	32.2	44.6	45.4	36.8	27.5	29.6	30.6	29.3	27.9	28.9	29.8	36.8	35.5	30.4	30.5
	SEM	10.3	13.5	11.0	18.6	18.6	9.4	9.0	6.0	9.0	9.8	6.8	4.6	5.4	13.3	11.3	4.2	4.6
K-R 20		.915	.910	.884	.826	.833	.935	.894	.959	.914	.889	.940	.975	.967	.870	.898	.981	.977

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 13 Grade 7		Reading	Written Expression	Conventions of Writing				Vocabulary	**ELA Total**	Mathematics			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
				Spelling	Capitalization	Punctuation	Conventions of Writing Total			Mathematics	Computation	**Math Total**						
		R	**WE**	**SP**	**CP**	**PC**	**CW**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**45**	**45**	**34**	**27**	**27**		**41**		**70**	**31**				**41**	**41**		
Fall																		
	Mean	29.2	26.7	19.2	13.5	12.4	–	22.3	–	39.7	16.7	–	–	–	22.5	23.3	–	–
RS	SD	9.2	9.0	7.1	5.3	5.4	–	7.9	–	13.7	6.9	–	–	–	8.3	8.4	–	–
	SEM	2.8	2.9	2.5	2.3	2.3	–	2.8	–	3.6	2.4	–	–	–	2.8	2.9	–	–
	Mean	231.3	234.7	233.5	235.4	236.8	232.9	231.2	232.7	232.8	231.8	232.5	232.6	232.8	233.9	233.2	232.9	233.0
SS	SD	36.3	46.3	32.8	45.9	47.0	38.2	28.2	30.4	31.7	30.1	28.5	29.9	30.8	37.8	36.4	31.4	31.8
	SEM	10.9	14.9	11.6	19.8	20.0	10.0	10.1	6.5	8.4	10.7	6.6	4.7	5.3	13.0	12.4	4.3	4.6
K-R 20		.910	.897	.875	.814	.818	.931	.871	.954	.930	.874	.946	.976	.970	.882	.885	.981	.979
Spring																		
	Mean	30.8	27.9	20.6	14.2	13.2	–	24.3	–	42.6	18.8	–	–	–	24.1	24.9	–	–
RS	SD	9.3	9.2	7.1	5.6	5.7	–	8.1	–	14.3	7.4	–	–	–	8.5	8.7	–	–
	SEM	2.7	2.8	2.5	2.3	2.3	–	2.8	–	3.5	2.3	–	–	–	2.8	2.8	–	–
	Mean	238.4	241.6	240.9	242.5	243.6	239.9	238.1	239.7	240.4	240.6	240.5	240.1	240.0	241.7	240.7	240.4	240.4
SS	SD	38.6	48.8	34.0	48.1	49.2	40.0	29.0	32.1	33.9	33.5	30.9	31.8	32.9	39.9	39.0	33.2	33.6
	SEM	11.1	15.0	11.8	19.4	19.6	9.9	9.9	6.6	8.4	10.6	6.6	4.7	5.3	12.9	12.5	4.3	4.7
K-R 20		.917	.906	.880	.838	.841	.939	.885	.958	.939	.899	.954	.978	.974	.895	.898	.983	.981

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 14 Grade 8		Reading	Written Expression	Conventions of Writing				Vocabulary	**ELA Total**	Mathematics			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
				Spelling	Capitalization	Punctuation	Conventions of Writing Total			Mathematics	Computation	**Math Total**						
		R	**WE**	**SP**	**CP**	**PC**	**CW**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**46**	**48**	**35**	**29**	**29**		**42**		**75**	**32**				**43**	**43**		
Fall																		
	Mean	29.8	28.3	19.0	15.3	13.9	–	22.8	–	42.9	18.3	–	–	–	23.0	24.8	–	–
RS	SD	9.6	10.3	7.3	5.8	6.2	–	8.4	–	14.3	7.0	–	–	–	7.9	9.0	–	–
	SEM	2.8	3.0	2.6	2.4	2.4	–	2.8	–	3.7	2.4	–	–	–	2.9	2.8	–	–
	Mean	242.3	245.2	244.4	246.0	247.2	243.4	241.9	243.4	244.2	243.9	244.1	243.8	243.8	245.0	244.2	244.0	244.1
SS	SD	39.5	50.2	34.5	49.0	50.0	41.0	29.7	32.8	34.5	34.1	31.6	32.6	33.5	40.6	39.8	33.9	34.4
	SEM	11.6	14.5	12.4	20.2	19.0	10.0	9.9	6.6	8.9	11.7	7.1	4.8	5.5	14.9	12.5	4.6	4.9
K-R 20		.913	.917	.870	.830	.856	.940	.889	.960	.934	.882	.950	.978	.973	.865	.901	.982	.980
Spring																		
	Mean	31.2	29.5	20.5	16.1	14.5	–	24.7	–	45.4	19.6	–	–	–	24.3	26.0	–	–
RS	SD	9.7	10.5	7.3	5.9	6.4	–	8.8	–	14.7	7.4	–	–	–	8.1	9.4	–	–
	SEM	2.7	2.9	2.6	2.3	2.3	–	2.8	–	3.6	2.3	–	–	–	2.9	2.8	–	–
	Mean	248.9	251.5	251.2	251.7	252.4	249.2	248.7	249.8	250.7	251.3	250.9	250.3	250.2	251.5	250.6	250.6	250.5
SS	SD	41.4	52.6	35.6	50.5	51.6	42.6	30.9	34.1	36.1	36.8	33.2	33.6	34.5	42.4	42.1	35.2	35.6
	SEM	11.7	14.6	12.5	19.9	18.7	9.9	9.7	6.6	8.9	11.7	7.1	4.9	5.5	15.0	12.5	4.6	4.9
K-R 20		.920	.923	.876	.845	.868	.946	.902	.962	.939	.899	.954	.979	.974	.875	.912	.983	.981

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 15 Grade 9		Reading	Written Expression	Vocabulary	**ELA Total**	**Mathematics**			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
						Mathematics	Computation	**Math Total**						
		R	**WE**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**40**	**54**	**40**		**40**	**30**				**48**	**50**		
Fall														
	Mean	22.1	27.8	19.3	–	16.2	13.2	–	–	–	20.5	19.2	–	–
RS	SD	9.1	11.5	9.2	–	7.9	5.8	–	–	–	8.8	8.7	–	–
	SEM	2.7	3.2	2.8	–	2.7	2.4	–	–	–	3.1	3.2	–	–
	Mean	252.4	254.7	251.8	253.5	254.0	254.4	254.1	253.8	253.7	254.3	253.6	253.8	253.8
SS	SD	42.4	43.0	31.4	34.7	36.6	37.5	33.8	34.0	34.8	42.6	42.7	35.5	36.0
	SEM	12.9	12.1	9.3	7.5	12.6	15.5	9.9	6.2	7.3	15.3	15.7	5.5	6.1
K-R 20		.913	.921	.908	.953	.882	.828	.915	.967	.955	.871	.866	.976	.971
Spring														
	Mean	23.4	29.1	21.2	–	17.5	14.1	–	–	–	21.7	20.6	–	–
RS	SD	9.4	11.6	9.6	–	8.4	6.2	–	–	–	9.2	9.4	–	–
	SEM	2.7	3.2	2.7	–	2.7	2.4	–	–	–	3.1	3.2	–	–
	Mean	258.8	260.2	258.2	259.4	259.9	259.6	259.8	259.6	259.7	260.4	259.6	259.7	259.8
SS	SD	44.4	43.3	32.7	35.8	38.0	39.0	34.9	34.5	35.6	43.5	43.7	36.5	36.8
	SEM	12.5	12.0	9.3	7.5	12.2	15.0	9.5	6.1	7.1	14.9	14.8	5.4	5.9
K-R 20		.920	.923	.920	.956	.897	.852	.925	.969	.960	.883	.886	.978	.974

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 16 Grade 10		Reading	Written Expression	Vocabulary	**ELA Total**	**Mathematics**			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
						Mathematics	Computation	**Math Total**						
		R	**WE**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**40**	**54**	**40**		**40**	**30**				**48**	**50**		
Fall														
	Mean	21.8	27.9	18.7	–	16.4	12.7	–	–	–	21.2	20.4	–	–
RS	SD	9.7	11.2	9.2	–	7.9	5.9	–	–	–	9.4	9.4	–	–
	SEM	2.7	3.3	2.7	–	2.7	2.4	–	–	–	3.1	3.2	–	–
	Mean	261.7	263.0	260.6	262.2	262.6	262.7	262.6	262.4	262.4	262.9	262.1	262.4	262.4
SS	SD	44.9	44.0	33.0	36.0	38.5	39.3	35.4	35.1	36.0	44.0	44.2	36.9	37.0
	SEM	13.5	12.9	9.3	8.0	13.3	16.0	10.4	6.5	7.8	14.5	14.7	5.6	6.2
K-R 20		.920	.914	.910	.951	.881	.834	.914	.965	.954	.891	.889	.977	.972
Spring														
	Mean	22.8	29.1	20.2	–	17.4	13.4	–	–	–	22.2	21.4	–	–
RS	SD	9.9	11.4	9.6	–	8.3	6.2	–	–	–	9.9	10.1	–	–
	SEM	2.7	3.3	2.7	–	2.7	2.4	–	–	–	3.1	3.1	–	–
	Mean	266.5	267.7	265.9	267.0	267.3	266.9	267.2	267.1	267.2	267.5	266.9	267.1	267.2
SS	SD	46.2	45.1	34.1	36.8	39.5	40.5	36.5	36.0	36.6	45.3	45.5	37.7	37.6
	SEM	13.1	12.8	9.3	7.9	13.0	15.5	10.1	6.4	7.6	14.2	14.2	5.4	6.1
K-R 20		.925	.919	.920	.954	.892	.853	.923	.968	.957	.902	.903	.979	.974

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 17/18 Grade 11		Reading	Written Expression	Vocabulary	**ELA Total**	Mathematics			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
						Mathematics								
		R	**WE**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**40**	**54**	**40**		**40**	**30**				**48**	**50**		
Fall														
	Mean	23.7	30.0	20.9	–	15.1	15.2	–	–	–	21.1	22.2	–	–
RS	SD	9.9	12.0	9.5	–	8.4	6.7	–	–	–	9.2	10.1	–	–
	SEM	2.6	3.2	2.7	–	2.7	2.4	–	–	–	3.1	3.1	–	–
	Mean	268.8	269.9	268.1	269.2	269.7	269.5	269.6	269.4	269.5	269.7	269.1	269.4	269.4
SS	SD	46.4	45.5	34.2	37.2	39.7	41.0	37.0	36.2	37.1	45.8	45.9	38.2	37.9
	SEM	13.2	12.2	9.1	7.7	12.9	14.6	9.9	6.2	7.5	15.4	14.3	5.5	6.1
K-R 20		.929	.928	.919	.958	.895	.874	.929	.970	.959	.887	.903	.980	.974
Spring														
	Mean	24.5	31.2	22.2	–	16.2	15.8	–	–	–	22.0	23.2	–	–
RS	SD	9.9	12.2	9.8	–	9.1	6.9	–	–	–	9.6	10.5	–	–
	SEM	2.6	3.2	2.7	–	2.7	2.4	–	–	–	3.1	3.1	–	–
	Mean	273.0	274.3	272.6	273.6	273.9	273.3	273.7	273.6	273.7	273.8	273.3	273.6	273.7
SS	SD	47.3	46.7	35.3	38.2	41.1	42.1	37.8	36.8	37.8	46.9	46.8	38.7	38.8
	SEM	12.9	12.1	9.2	7.5	12.3	14.4	9.5	6.1	7.2	15.1	14.0	5.3	5.9
K-R 20		.932	.933	.926	.961	.910	.883	.936	.973	.963	.896	.911	.981	.977

Continued on next page…

Table 11 (*continued*): Means, Standard Deviations (SD), Reliability Coefficients (K-R 20), and Standard Errors of Measurement (SEM) for the Weighted Sample, Grades K–12
***Iowa Assessments* Form E**

Level 17/18 Grade 12		Reading	Written Expression	Vocabulary	**ELA Total**	**Mathematics**			**Core Composite**	Core Composite with ET and M	Science	Social Studies	**Complete Composite**	Complete Composite with ET and M
						Mathematics	Computation	**Math Total**						
		R	**WE**	**V**	**ET**	**M**	**MC**	**MT**	**CT**	**CT-**	**SC**	**SS**	**CC**	**CC-**
Number of Items		**40**	**54**	**40**		**40**	**30**				**48**	**50**		
Fall														
	Mean	25.0	31.6	22.5	–	16.5	16.0	–	–	–	22.3	23.7	–	–
RS	SD	9.9	12.2	9.8	–	9.3	7.0	–	–	–	9.7	10.7	–	–
	SEM	2.6	3.2	2.6	–	2.7	2.4	–	–	–	3.1	3.1	–	–
	Mean	274.6	276.1	273.9	275.2	275.6	275.1	275.4	275.3	275.4	274.9	275.1	275.2	275.3
SS	SD	47.3	46.8	35.3	38.2	41.4	42.5	38.1	37.1	37.8	47.1	47.0	39.0	39.2
	SEM	12.8	12.1	9.2	7.5	12.1	14.3	9.4	6.0	7.1	15.0	13.7	5.3	5.8
K-R 20		.932	.933	.927	.961	.914	.887	.939	.974	.964	.898	.915	.982	.978
Spring														
	Mean	25.6	32.5	23.6	–	17.3	16.6	–	–	–	23.1	24.7	–	–
RS	SD	9.9	12.4	9.9	–	9.7	7.2	–	–	–	10.0	11.2	–	–
	SEM	2.6	3.1	2.6	–	2.7	2.3	–	–	–	3.1	3.1	–	–
	Mean	278.2	279.6	277.6	278.8	278.8	278.7	278.8	178.8	278.8	278.3	279.1	278.8	278.8
SS	SD	48.1	47.7	36.2	38.8	42.3	43.4	38.7	37.8	38.2	47.7	47.9	39.4	39.5
	SEM	12.5	12.0	9.3	7.4	11.8	14.1	9.2	5.9	7.0	14.8	13.3	5.2	5.7
K-R 20		.934	.937	.932	.963	.922	.894	.944	.976	.967	.904	.923	.983	.979

Sources of Variation in Measurement

Further investigation of sources of variation that might affect scores on large-scale assessments was provided in two studies of reliability based on test administrations from multiple occasions. The first used data from the 2010 equating of Form E of the *Iowa Assessments* and Form A of the *ITBS*/*ITED.* The second used data from a 2011–12 comparability study involving Levels 5/6 through 17/18 in grades kindergarten through 11.

As previously described, Form E of the *Iowa Assessments* and Form A of the *ITBS*/*ITED* were administered to a large national sample of schools that were selected to be representative with respect to variability in achievement. The matched records from this study made an analysis of relative contributions of various sources of measurement error across tests, grades, and schools possible. Results are reported for the Reading and Mathematics tests.

In addition to alternate-forms reliability coefficients, three other "within-forms" reliability coefficients were computed.

- K-R20 reliability coefficients were calculated from the item-response records.
- Split-halves coefficients were computed by correlating raw scores from odd-numbered versus even-numbered items. Full-test reliabilities were estimated using the Spearman-Brown formula.
- Split-halves coefficients were computed by correlating raw scores from items in the separately timed Part 1 and Part 2 of the Reading and Mathematics tests. Again, full-test reliabilities were estimated using the Spearman-Brown formula.

Table 12 presents the results of the analysis of the within-forms and between-forms estimates of reliability. Differences between within-forms estimates obtained in the same testing session ($K\text{-}R_{20}$, SH_{OE} and $SH_{pt1/pt2}$) and alternate-forms estimates obtained a week or two apart ($A\text{-}F$) constitute the best evidence on the effects of changes in student motivation and behavior across several days.

Although the median reliability coefficients for the same-day estimates are quite similar (and expected to be so in that $K\text{-}R_{20}$ is the theoretical average of all possible split-half coefficients), there are small differences between same-day and different-day estimates.

Table 12: Reliability Coefficients based on Split-Halves from Odd-Even (SH_{OE}) and Timed Parts ($SH_{pt1/pt2}$) and on Alternate Forms (*A-F*)
***Iowa Assessments* Form E 2010 National Standardization**

Grade	Level	Reading				Mathematics			
		K-R20	SHOE	$SH_{pt1/pt2}$	A-F	K-R20	SHOE	$SH_{pt1/pt2}$	A-F
3	**9**	.90	.90	.85	.84	.86	.88	.84	.83
4	**10**	.90	.91	.87	.82	.88	.89	.86	.84
5	**11**	.91	.91	.87	.82	.89	.90	.86	.84
6	**12**	.91	.92	.88	.83	.91	.91	.88	.85
7	**13**	.91	.91	.88	.80	.93	.93	.91	.86
8	**14**	.91	.92	.89	.84	.93	.94	.91	.89
9	**15**	.91	.92	.80*	.72	.88	.89	.80	.75
10	**16**	.92	.93	.82*	.76	.88	.89	.79*	.78
11	**17**	.93	.94	.84*	.58	.90	.89	.83*	.51
Median		.91	.92	.86	.85	.89	.89	.85	.84

Another study of sources of variation in measurement was completed during the 2011–2012 comparability study of paper-based and computer-based administrations of the *Iowa Assessments* Form E. In this study, the same students took Form E in both administration modes. The order of testing modes was counter-balanced, and an interval of between one and two weeks separated the two administrations. Correlations between scores in different modes can be interpreted as estimates of test-retest reliability. While the mode of administration does represent an additional source of variation in these scores, high correlations constitute evidence that the combined effects of temporal changes in examinees and administrative conditions are small. These correlations are reported in Table 13. The median test-retest correlations range from .73 in Mathematics and Social Studies to .80 in Reading and Written Expression. The values are predictably lower than internal-consistency reliability estimates reported previously and somewhat lower than the alternate-forms coefficients reported in Table 12. This is probably due to the presence of a small amount of variation in scores due to the mode of administration—paper-based versus computer-based.

Table 13: *Iowa Assessments*, Form E, Estimates of Test-Retest Reliability

Level (Grade)	N	R	WE	M	SS	SC
5/6 (1)	1192	0.86	0.68	0.66	–	–
7 (2)	1059	0.82	0.76	0.76	0.63	0.65
8 (2)	1073	0.87	0.74	0.74	0.69	0.71
9 (3)	253	0.68	0.80	0.71	0.70	0.77
10 (4)	249	0.81	0.84	0.86	0.73	0.77
11 (5)	254	0.82	0.79	0.68	0.72	0.67
12 (6)	329	0.71	0.90	0.72	0.85	0.84
13 (7)	306	0.80	0.92	0.83	0.90	0.89
14 (8)	314	0.79	0.82	0.73	0.80	0.82
15 (9)	282	0.73	0.65	0.77	0.74	0.75
16 (10)	292	0.68	0.59	0.57	0.81	0.67
17 (11/12)	372	0.77	0.74	0.73	0.66	0.65
Average	–	.80	.80	.73	.73	.75

Note: Correlations include occasion and mode of administration variance.

Standard Errors of Measurement for Selected Score Levels

Examinee-level errors of measurement based on a single test administration, conditional standard errors of measurement (CSEMs) were estimated using several procedures identified by previous studies to yield similar results (Qualls-Payne, 1992; Feldt & Qualls, 1998).

The results in Table 14 were obtained using a method developed by Brennan and Lee (1997) for smoothing a plot of conditional standard errors of scaled scores based on the binomial error model. In addition to this method, an approach developed by Feldt and Qualls (1998) and another based on bootstrap techniques were used at selected test levels. Because the results of all three methods agreed closely and generally matched the patterns of varying SEMs by score level found with previous editions of the *ITBS* and *ITED*, only the results of the Brennan and Lee method are provided.

Table 14: Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 5/6	Reading	Language	Vocabulary	Word Analysis	Listening	Mathematics	Reading Words	Reading Comprehension
NSS Range	**R**	**L**	**V**	**WA**	**Li**	**M**	**RW**	**RC**
90-99	–	2.68	2.39	2.59	4.63	2.50	–	–
100-109	3.00	5.39	3.65	4.78	6.18	4.30	–	5.67
110-119	4.75	8.91	5.21	4.13	5.48	4.97	7.61	7.53
120-129	5.28	10.77	6.94	5.14	6.11	4.33	7.23	8.53
130-139	3.43	10.77	6.92	6.91	6.74	5.17	3.85	5.95
140-149	2.25	11.24	7.54	11.19	7.56	5.29	4.97	3.71
150-159	5.94	12.22	7.23	11.97	8.00	6.06	7.67	8.00
160-169	–	12.54	6.13	10.00	10.26	6.89	–	–
170-179	–	12.20	4.25	–	7.75	–	–	–
180-189	–	8.50	–	–	–	–	–	–

Level 7	Reading	Vocabulary	Language	Word Analysis	Listening	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**L**	**WA**	**Li**	**M**	**MC**	**SS**	**SC**
100-109	–	4.24	4.07	2.39	3.51	2.42	–	2.82	2.23
110-119	3.24	7.16	6.61	5.44	6.01	5.53	6.11	4.86	5.80
120-129	6.57	9.18	7.09	6.69	6.40	7.26	7.81	5.70	7.72
130-139	5.44	8.34	6.37	7.69	6.90	7.78	4.80	7.78	10.52
140-149	3.48	5.34	5.70	8.01	6.86	7.50	3.68	9.55	12.04
150-159	4.51	5.76	4.55	9.14	7.42	6.99	4.64	10.63	12.73
160-169	8.60	7.98	5.45	10.93	10.16	6.38	4.49	11.69	13.66
170-179	8.64	10.94	5.83	10.30	11.09	6.47	–	12.24	12.80
180-189	7.00	8.75	4.00	8.00	9.00	4.89	–	9.67	11.39
190-199	–	–	–	–	–	–	–	–	8.75

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 8	Reading	Vocabulary	Language	Word Analysis	Listening	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**L**	**WA**	**Li**	**M**	**MC**	**SS**	**SC**
100-109	–	3.86	2.40	2.00	3.50	1.87	–	3.25	2.00
110-119	3.61	6.80	4.76	5.21	6.48	4.52	4.97	5.53	4.30
120-129	7.52	9.55	7.62	7.63	7.98	7.04	8.53	7.11	6.86
130-139	7.46	10.78	8.13	8.06	6.90	7.32	8.84	6.18	7.22
140-149	5.36	10.51	6.98	7.82	7.13	6.31	4.86	6.72	8.54
150-159	4.78	9.37	5.72	8.33	8.36	6.78	5.76	9.15	11.16
160-169	6.18	9.39	4.86	10.76	8.16	7.21	8.18	10.95	12.85
170-179	8.29	10.78	7.25	13.35	9.90	7.69	10.30	12.16	14.16
180-189	11.30	10.94	9.50	15.23	12.73	8.06	12.20	11.73	15.91
190-199	12.73	11.88	9.76	17.33	12.76	8.36	9.92	12.22	16.31
200-209	11.16	10.53	8.38	15.59	10.25	7.01	–	11.11	15.02
210-219	8.50	8.25	6.40	12.00	–	4.80	–	8.25	13.05
220-229	–	–	–	–	–	–	–	–	9.50

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 9	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
120-129	3.81	3.25	–	–	–	4.12	2.80	–	–	4.25
130-139	5.73	5.88	6.75	7.14	9.18	6.35	4.52	5.75	5.70	5.86
140-149	7.19	10.48	9.15	10.96	12.77	8.89	5.24	7.63	6.99	6.57
150-159	8.28	12.97	9.32	14.59	14.35	11.49	6.68	8.25	7.12	7.78
160-169	7.97	12.60	8.21	15.50	15.59	11.26	7.51	8.29	8.60	9.89
170-179	6.92	9.10	8.08	14.19	15.30	8.64	7.74	8.92	9.05	10.97
180-189	6.96	6.39	8.36	10.39	13.12	6.75	7.85	7.37	9.03	10.67
190-199	8.10	6.16	8.24	9.73	11.98	8.31	7.96	7.01	9.26	9.86
200-209	10.77	7.00	9.86	14.75	15.13	11.70	9.21	8.25	11.38	12.12
210-219	12.66	9.15	13.08	19.78	18.48	14.17	10.28	10.82	11.94	15.25
220-229	13.90	14.27	15.99	22.13	20.12	16.54	10.27	10.82	11.62	16.32
230-239	14.60	13.37	17.20	23.15	19.78	16.33	9.44	8.75	11.79	15.12
240-249	14.58	11.00	15.81	–	18.11	15.84	8.71	–	13.54	12.58
250-259	11.87	–	12.50	22.24	16.20	15.83	7.65	–	–	12.49
260-269	9.00	–	–	21.74	14.42	13.67	5.55	–	10.25	9.00
270-279	–	–	–	–	13.80	10.00	–	–	–	–
280-289	–	–	–	16.67	10.67	–	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 10	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
120-129	3.81	3.25	–	–	–	4.12	2.80	–	–	4.25
130-139	5.73	5.88	6.75	7.14	9.18	6.35	4.52	5.75	5.70	5.86
140-149	7.19	10.48	9.15	10.96	12.77	8.89	5.24	7.63	6.99	6.57
150-159	8.28	12.97	9.32	14.59	14.35	11.49	6.68	8.25	7.12	7.78
160-169	7.97	12.60	8.21	15.50	15.59	11.26	7.51	8.29	8.60	9.89
170-179	6.92	9.10	8.08	14.19	15.30	8.64	7.74	8.92	9.05	10.97
180-189	6.96	6.39	8.36	10.39	13.12	6.75	7.85	7.37	9.03	10.67
190-199	8.10	6.16	8.24	9.73	11.98	8.31	7.96	7.01	9.26	9.86
200-209	10.77	7.00	9.86	14.75	15.13	11.70	9.21	8.25	11.38	12.12
210-219	12.66	9.15	13.08	19.78	18.48	14.17	10.28	10.82	11.94	15.25
220-229	13.90	14.27	15.99	22.13	20.12	16.54	10.27	10.82	11.62	16.32
230-239	14.60	13.37	17.20	23.15	19.78	16.33	9.44	8.75	11.79	15.12
240-249	14.58	11.00	15.81	–	18.11	15.84	8.71	–	13.54	12.58
250-259	11.87	–	12.50	22.24	16.20	15.83	7.65	–	–	12.49
260-269	9.00	–	–	21.74	14.42	13.67	5.55	–	10.25	9.00
270-279	–	–	–	–	13.80	10.00	–	–	–	–
280-289	–	–	–	16.67	10.67	–	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 11	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
120-129	–	–	–	–	–	3.48	–	–	–	–
130-139	5.07	4.62	6.50	6.00	7.33	5.78	3.28	6.50	6.25	5.18
140-149	7.03	7.76	8.34	9.11	10.63	9.94	5.56	8.33	8.53	6.61
150-159	7.32	12.39	9.17	13.43	13.52	12.48	7.45	9.31	8.66	7.11
160-169	7.89	14.11	9.61	16.25	15.75	13.40	8.27	9.68	7.11	9.50
170-179	9.08	13.10	10.03	17.06	18.04	12.57	8.63	9.97	9.14	10.79
180-189	8.97	8.79	9.37	15.83	17.10	8.62	8.97	9.28	11.43	12.55
190-199	8.76	7.48	8.79	13.95	16.35	7.67	8.98	7.90	12.31	12.59
200-209	8.81	7.56	9.24	13.76	15.50	9.59	8.74	7.68	12.57	11.87
210-219	9.75	6.64	10.35	15.19	18.11	12.31	9.07	9.35	9.75	13.84
220-229	11.51	7.32	12.63	18.17	21.59	15.30	10.45	10.53	8.76	16.39
230-239	12.88	11.18	15.20	21.82	23.02	17.62	11.37	12.28	11.86	17.34
240-249	14.61	13.94	17.36	26.16	23.00	18.58	10.51	12.93	13.27	16.85
250-259	16.38	11.50	17.83	26.55	21.61	19.59	8.73	9.75	12.65	14.89
260-269	15.21	–	15.84	22.86	18.42	20.66	8.23	–	15.00	12.62
270-279	13.27	–	–	18.18	14.73	19.04	6.52	–	14.22	11.58
280-289	10.20	–	12.00	14.07	11.21	16.29	4.67	–	11.25	9.69
290-299	–	–	–	12.79	12.10	–	–	–	–	–
300-309	–	–	–	10.33	10.00	12.20	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 12	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
120-129	–	–	–	–	–	2.40	–	–	–	–
130-139	4.31	5.25	–	6.25	6.50	5.96	3.65	–	4.50	5.50
140-149	6.35	7.84	7.00	9.96	9.82	10.58	5.35	7.43	7.35	7.42
150-159	7.59	9.06	9.90	12.78	12.78	13.36	6.86	10.03	9.16	7.88
160-169	8.83	9.40	10.44	14.97	16.19	13.92	7.33	10.90	10.62	8.52
170-179	9.30	9.34	10.00	17.33	18.71	12.93	8.06	11.84	11.92	12.16
180-189	9.27	10.60	9.58	18.65	20.23	11.34	8.50	11.91	12.54	14.76
190-199	9.42	11.35	9.26	19.19	21.23	10.03	8.32	11.16	11.95	15.28
200-209	9.48	9.99	10.03	19.08	21.25	11.25	8.93	9.40	10.15	14.06
210-219	10.42	9.05	10.60	19.55	21.53	13.91	9.87	9.19	10.50	12.46
220-229	12.51	9.93	10.79	21.25	22.73	15.48	9.84	9.89	11.95	13.37
230-239	14.13	9.48	12.72	23.22	24.06	17.31	10.47	10.64	12.54	15.47
240-249	13.85	8.21	14.65	24.17	24.17	19.22	10.23	11.45	13.08	17.04
250-259	15.26	9.69	15.66	25.31	24.09	19.91	10.46	14.29	13.49	17.54
260-269	17.44	14.05	16.47	24.99	22.61	19.69	10.04	13.90	12.95	17.04
270-279	16.19	13.12	16.75	23.29	20.62	18.98	9.67	–	12.89	15.01
280-289	13.98	10.75	14.75	20.70	17.44	18.11	8.99	11.25	16.58	11.60
290-299	–	–	11.25	17.35	14.97	18.06	6.94	–	15.07	9.92
300-309	11.00	–	–	14.98	13.22	16.42	4.67	–	12.00	9.49
310-319	–	–	–	11.91	10.52	13.74	–	–	–	7.50
320-329	–	–	–	9.00	7.75	10.00	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 13	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
120-129	–	–	–	–	–	2.60	–	–	–	–
130-139	4.40	4.80	–	5.50	7.00	4.63	3.17	–	5.20	–
140-149	6.18	7.12	7.75	9.01	9.98	7.80	4.85	7.50	6.98	6.19
150-159	6.89	9.22	10.03	12.23	12.76	11.77	6.61	10.29	8.70	8.31
160-169	6.90	11.39	11.16	16.53	17.25	14.24	7.77	12.78	9.31	9.22
170-179	8.84	13.46	10.32	20.11	20.91	15.73	8.69	14.22	10.27	11.06
180-189	10.41	15.03	10.57	21.76	–	15.10	9.28	14.75	11.65	12.57
190-199	11.93	15.28	11.84	23.30	22.82	13.69	9.54	14.03	13.05	13.81
200-209	11.87	13.75	12.12	23.44	23.78	12.88	9.75	13.19	12.91	12.37
210-219	11.79	12.10	12.07	22.98	23.36	12.71	9.74	11.79	12.52	11.52
220-229	11.64	11.20	11.83	22.36	22.98	14.23	9.05	10.82	12.38	12.68
230-239	12.02	10.98	12.12	22.93	22.80	16.13	8.03	10.62	12.91	13.42
240-249	12.38	9.62	13.06	23.53	23.37	19.01	8.93	10.81	13.94	15.12
250-259	12.84	8.08	14.15	24.28	23.81	20.37	9.96	10.93	14.35	15.61
260-269	14.86	9.26	14.23	23.87	23.68	21.00	10.00	10.73	13.96	16.36
270-279	15.99	11.13	14.32	23.30	22.98	21.15	8.63	11.70	13.95	17.08
280-289	17.86	12.52	13.82	20.55	21.48	20.69	7.70	13.79	13.44	15.74
290-299	16.68	10.71	13.71	17.79	18.54	19.45	8.63	–	15.59	14.33
300-309	14.91	8.00	11.98	15.04	16.63	16.98	7.98	10.50	15.43	13.10
310-319	11.60	–	9.00	12.30	14.38	15.72	5.67	–	13.73	11.27
320-329	–	–	–	12.06	12.23	14.97	–	–	10.60	7.80
330-339	–	–	–	9.00	10.26	11.88	–	–	–	–
340-349	–	–	–	–	7.00	8.80	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 14	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
130-139	4.00	–	–	–	–	4.09	–	–	–	–
140-149	5.82	6.42	7.25	6.24	8.11	6.95	3.79	6.50	5.69	6.19
150-159	6.69	9.64	9.75	9.81	12.52	10.94	6.42	8.90	8.87	7.73
160-169	7.32	11.89	12.25	14.51	16.06	14.19	7.88	11.32	11.13	8.47
170-179	10.68	13.97	13.30	17.59	18.98	15.95	8.77	13.35	13.04	9.67
180-189	12.76	15.14	13.98	19.60	21.83	16.27	8.84	15.05	14.25	13.40
190-199	13.73	15.61	14.14	21.13	23.06	15.79	9.65	16.06	14.68	16.16
200-209	13.30	15.55	13.51	22.14	23.68	14.54	9.70	16.87	14.70	17.91
210-219	12.42	14.42	12.69	21.65	23.24	13.27	9.30	16.99	13.37	17.44
220-229	11.75	13.00	12.56	21.80	21.27	13.12	9.13	15.13	11.85	16.01
230-239	12.13	11.85	12.82	22.37	19.79	13.45	9.68	13.40	11.71	15.33
240-249	12.14	10.91	12.67	23.28	20.26	14.05	10.64	12.34	12.63	15.07
250-259	12.28	9.20	12.90	24.41	21.45	15.79	10.50	11.94	13.69	16.29
260-269	12.48	7.67	13.61	25.71	21.80	17.11	10.00	11.16	14.68	17.32
270-279	14.06	8.41	14.66	25.18	22.04	18.99	9.64	9.86	15.31	17.70
280-289	15.10	10.32	14.55	24.09	22.16	20.60	9.42	12.22	15.52	17.86
290-299	18.60	13.13	13.95	22.05	20.57	20.73	10.31	15.36	15.09	17.63
300-309	17.41	11.41	13.34	19.54	18.78	19.57	9.71	14.19	14.66	16.87
310-319	15.38	8.80	13.44	17.02	17.59	18.53	9.57	11.50	15.39	14.79
320-329	12.00	–	11.70	14.51	15.56	17.66	8.09	–	13.42	12.30
330-339	–	–	9.00	13.91	14.17	17.03	5.18	–	–	10.73
340-349	–	–	–	11.94	12.79	14.77	–	–	9.20	7.08
350-359	–	–	–	9.25	8.75	12.36	–	–	–	–
360-369	–	–	–	–	–	9.20	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 14	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
130-139	4.00	–	–	–	–	4.09	–	–	–	–
140-149	5.82	6.42	7.25	6.24	8.11	6.95	3.79	6.50	5.69	6.19
150-159	6.69	9.64	9.75	9.81	12.52	10.94	6.42	8.90	8.87	7.73
160-169	7.32	11.89	12.25	14.51	16.06	14.19	7.88	11.32	11.13	8.47
170-179	10.68	13.97	13.30	17.59	18.98	15.95	8.77	13.35	13.04	9.67
180-189	12.76	15.14	13.98	19.60	21.83	16.27	8.84	15.05	14.25	13.40
190-199	13.73	15.61	14.14	21.13	23.06	15.79	9.65	16.06	14.68	16.16
200-209	13.30	15.55	13.51	22.14	23.68	14.54	9.70	16.87	14.70	17.91
210-219	12.42	14.42	12.69	21.65	23.24	13.27	9.30	16.99	13.37	17.44
220-229	11.75	13.00	12.56	21.80	21.27	13.12	9.13	15.13	11.85	16.01
230-239	12.13	11.85	12.82	22.37	19.79	13.45	9.68	13.40	11.71	15.33
240-249	12.14	10.91	12.67	23.28	20.26	14.05	10.64	12.34	12.63	15.07
250-259	12.28	9.20	12.90	24.41	21.45	15.79	10.50	11.94	13.69	16.29
260-269	12.48	7.67	13.61	25.71	21.80	17.11	10.00	11.16	14.68	17.32
270-279	14.06	8.41	14.66	25.18	22.04	18.99	9.64	9.86	15.31	17.70
280-289	15.10	10.32	14.55	24.09	22.16	20.60	9.42	12.22	15.52	17.86
290-299	18.60	13.13	13.95	22.05	20.57	20.73	10.31	15.36	15.09	17.63
300-309	17.41	11.41	13.34	19.54	18.78	19.57	9.71	14.19	14.66	16.87
310-319	15.38	8.80	13.44	17.02	17.59	18.53	9.57	11.50	15.39	14.79
320-329	12.00	–	11.70	14.51	15.56	17.66	8.09	–	13.42	12.30
330-339	–	–	9.00	13.91	14.17	17.03	5.18	–	–	10.73
340-349	–	–	–	11.94	12.79	14.77	–	–	9.20	7.08
350-359	–	–	–	9.25	8.75	12.36	–	–	–	–
360-369	–	–	–	–	–	9.20	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 15	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
130-139	4.00	–	–	–	–	4.09	–	–	–	–
140-149	5.82	6.42	7.25	6.24	8.11	6.95	3.79	6.50	5.69	6.19
150-159	6.69	9.64	9.75	9.81	12.52	10.94	6.42	8.90	8.87	7.73
160-169	7.32	11.89	12.25	14.51	16.06	14.19	7.88	11.32	11.13	8.47
170-179	10.68	13.97	13.30	17.59	18.98	15.95	8.77	13.35	13.04	9.67
180-189	12.76	15.14	13.98	19.60	21.83	16.27	8.84	15.05	14.25	13.40
190-199	13.73	15.61	14.14	21.13	23.06	15.79	9.65	16.06	14.68	16.16
200-209	13.30	15.55	13.51	22.14	23.68	14.54	9.70	16.87	14.70	17.91
210-219	12.42	14.42	12.69	21.65	23.24	13.27	9.30	16.99	13.37	17.44
220-229	11.75	13.00	12.56	21.80	21.27	13.12	9.13	15.13	11.85	16.01
230-239	12.13	11.85	12.82	22.37	19.79	13.45	9.68	13.40	11.71	15.33
240-249	12.14	10.91	12.67	23.28	20.26	14.05	10.64	12.34	12.63	15.07
250-259	12.28	9.20	12.90	24.41	21.45	15.79	10.50	11.94	13.69	16.29
260-269	12.48	7.67	13.61	25.71	21.80	17.11	10.00	11.16	14.68	17.32
270-279	14.06	8.41	14.66	25.18	22.04	18.99	9.64	9.86	15.31	17.70
280-289	15.10	10.32	14.55	24.09	22.16	20.60	9.42	12.22	15.52	17.86
290-299	18.60	13.13	13.95	22.05	20.57	20.73	10.31	15.36	15.09	17.63
300-309	17.41	11.41	13.34	19.54	18.78	19.57	9.71	14.19	14.66	16.87
310-319	15.38	8.80	13.44	17.02	17.59	18.53	9.57	11.50	15.39	14.79
320-329	12.00	–	11.70	14.51	15.56	17.66	8.09	–	13.42	12.30
330-339	–	–	9.00	13.91	14.17	17.03	5.18	–	–	10.73
340-349	–	–	–	11.94	12.79	14.77	–	–	9.20	7.08
350-359	–	–	–	9.25	8.75	12.36	–	–	–	–
360-369	–	–	–	–	–	9.20	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 16	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
130-139	4.00	–	–	–	–	4.09	–	–	–	–
140-149	5.82	6.42	7.25	6.24	8.11	6.95	3.79	6.50	5.69	6.19
150-159	6.69	9.64	9.75	9.81	12.52	10.94	6.42	8.90	8.87	7.73
160-169	7.32	11.89	12.25	14.51	16.06	14.19	7.88	11.32	11.13	8.47
170-179	10.68	13.97	13.30	17.59	18.98	15.95	8.77	13.35	13.04	9.67
180-189	12.76	15.14	13.98	19.60	21.83	16.27	8.84	15.05	14.25	13.40
190-199	13.73	15.61	14.14	21.13	23.06	15.79	9.65	16.06	14.68	16.16
200-209	13.30	15.55	13.51	22.14	23.68	14.54	9.70	16.87	14.70	17.91
210-219	12.42	14.42	12.69	21.65	23.24	13.27	9.30	16.99	13.37	17.44
220-229	11.75	13.00	12.56	21.80	21.27	13.12	9.13	15.13	11.85	16.01
230-239	12.13	11.85	12.82	22.37	19.79	13.45	9.68	13.40	11.71	15.33
240-249	12.14	10.91	12.67	23.28	20.26	14.05	10.64	12.34	12.63	15.07
250-259	12.28	9.20	12.90	24.41	21.45	15.79	10.50	11.94	13.69	16.29
260-269	12.48	7.67	13.61	25.71	21.80	17.11	10.00	11.16	14.68	17.32
270-279	14.06	8.41	14.66	25.18	22.04	18.99	9.64	9.86	15.31	17.70
280-289	15.10	10.32	14.55	24.09	22.16	20.60	9.42	12.22	15.52	17.86
290-299	18.60	13.13	13.95	22.05	20.57	20.73	10.31	15.36	15.09	17.63
300-309	17.41	11.41	13.34	19.54	18.78	19.57	9.71	14.19	14.66	16.87
310-319	15.38	8.80	13.44	17.02	17.59	18.53	9.57	11.50	15.39	14.79
320-329	12.00	–	11.70	14.51	15.56	17.66	8.09	–	13.42	12.30
330-339	–	–	9.00	13.91	14.17	17.03	5.18	–	–	10.73
340-349	–	–	–	11.94	12.79	14.77	–	–	9.20	7.08
350-359	–	–	–	9.25	8.75	12.36	–	–	–	–
360-369	–	–	–	–	–	9.20	–	–	–	–

Table 14 (*continued*): Standard Errors of Measurement for Selected NSS Levels
***Iowa Assessments* Form E**

Level 17	Reading	Vocabulary	Spelling	Capitalization	Punctuation	Written Expression	Mathematics	Computation	Social Studies	Science
NSS Range	**R**	**V**	**SP**	**CP**	**PC**	**WE**	**M**	**MC**	**SS**	**SC**
130-139	4.00	–	–	–	–	4.09	–	–	–	–
140-149	5.82	6.42	7.25	6.24	8.11	6.95	3.79	6.50	5.69	6.19
150-159	6.69	9.64	9.75	9.81	12.52	10.94	6.42	8.90	8.87	7.73
160-169	7.32	11.89	12.25	14.51	16.06	14.19	7.88	11.32	11.13	8.47
170-179	10.68	13.97	13.30	17.59	18.98	15.95	8.77	13.35	13.04	9.67
180-189	12.76	15.14	13.98	19.60	21.83	16.27	8.84	15.05	14.25	13.40
190-199	13.73	15.61	14.14	21.13	23.06	15.79	9.65	16.06	14.68	16.16
200-209	13.30	15.55	13.51	22.14	23.68	14.54	9.70	16.87	14.70	17.91
210-219	12.42	14.42	12.69	21.65	23.24	13.27	9.30	16.99	13.37	17.44
220-229	11.75	13.00	12.56	21.80	21.27	13.12	9.13	15.13	11.85	16.01
230-239	12.13	11.85	12.82	22.37	19.79	13.45	9.68	13.40	11.71	15.33
240-249	12.14	10.91	12.67	23.28	20.26	14.05	10.64	12.34	12.63	15.07
250-259	12.28	9.20	12.90	24.41	21.45	15.79	10.50	11.94	13.69	16.29
260-269	12.48	7.67	13.61	25.71	21.80	17.11	10.00	11.16	14.68	17.32
270-279	14.06	8.41	14.66	25.18	22.04	18.99	9.64	9.86	15.31	17.70
280-289	15.10	10.32	14.55	24.09	22.16	20.60	9.42	12.22	15.52	17.86
290-299	18.60	13.13	13.95	22.05	20.57	20.73	10.31	15.36	15.09	17.63
300-309	17.41	11.41	13.34	19.54	18.78	19.57	9.71	14.19	14.66	16.87
310-319	15.38	8.80	13.44	17.02	17.59	18.53	9.57	11.50	15.39	14.79
320-329	12.00	–	11.70	14.51	15.56	17.66	8.09	–	13.42	12.30
330-339	–	–	9.00	13.91	14.17	17.03	5.18	–	–	10.73
340-349	–	–	–	11.94	12.79	14.77	–	–	9.20	7.08
350-359	–	–	–	9.25	8.75	12.36	–	–	–	–
360-369	–	–	–	–	–	9.20	–	–	–	–