Response to Claims Raised by DRC's Mathematics and Reading Alignment Study

The Iowa Department of Education released a report in October 2013 that was commissioned in March 2013 to align the *Iowa Assessments*, Form E, to the Iowa Core/Common Core State Standards in Reading and Mathematics. The study compared Form E to a set of standards that were not used to develop Form E. Because of flaws in the methodology and implementation of the alignment study, Iowa Testing Programs (ITP) has serious reservations about the integrity of the results and serious questions about the validity of any conclusions drawn by the authors, Data Recognition Corporation, of the report.

November 2013

Executive Summary

Purpose. The Iowa Department of Education released a report commissioned in March 2013 to evaluate the alignment of the *Iowa Assessments*, Form E, to the Iowa Core/Common Core State Standards in Reading and Mathematics. Form E was built to the Iowa Core as the standards were mandated in 2008, prior to the development and adoption of the Common Core State Standards (CCSS). The 2008 version of the Iowa Core was based on grade spans (primary, intermediate, middle and high school). Form E of the *Iowa Assessments* (and a parallel Form F) were built as a transition to serve as a bridge between the adoption of the Iowa Core in 2008 and the full implementation of the Iowa Core in 2015, while continuing to meet the reporting requirements of No Child Left Behind (NCLB). This transition was a necessary step for the state of Iowa to measure the Iowa Core of 2008 while reporting AYP and proficiency information to the federal government. This decision was discussed explicitly with the Iowa Department of Education as plans for the transition were made.

The commissioned report explicitly states that the *Iowa Assessments*, Form E, were not built to the standards that were used to complete the study. This document cites text excerpted from this report (pages 4 and 15) that refutes any findings in this report for purposes of alignment to the Common Core adopted by Iowa in 2010.

Methodology. The study used a judgmental method to determine the correspondence between individual questions and the Iowa Core/Common Core. After a training period, multiple judges attempted to match questions to specific standards at grade levels 3-8 and 11. In the end, judges achieved such poor levels of agreement that the only trustworthy conclusion supported by the data is that the training was inadequate to the task. In one case, 15 different judges matched the same question to 11 different standards in the Iowa/Common Core Standards in Mathematics, ranging across four grade levels.

The lack of agreement of the judges' ratings in the DRC report, acknowledged by its authors, precludes any absolute statements about alignment of the *Iowa Assessments*, Form E.

Implementation. New development underway since 2011 will support the implementation of the next generation of the *Iowa Assessments* built to the precise specifications of the Iowa Core.

As assessment providers to Iowa schools and to the Iowa Department of Education, Iowa Testing Programs is committed to ongoing research and development to measure the full range of standards deemed important to Iowa educators.

In response to the claims made in the report, this section addresses specific questions about the purpose, methodology and results of the study.

Background

Q: What is “Alignment” in Testing?

A: Alignment is the correspondence between test questions and the subject matter and standards the test is intended to measure.

Q: How can you tell is if a test is “aligned” as intended?

A: Test developers align questions to standards as the test is developed. After the test is developed, a judgmental process may be used to validate the test developers’ efforts.

Q: By “a judgmental process” do you mean that alignment is subjective?

A: Not necessarily. A sound approach to alignment requires extensive training of judges and regular evaluations of inter-judge agreement. These same procedures are used in large-scale assessment to train raters to score essays and to train judges that set passing scores on certification tests.

Q: Is it possible for a single test to be “fully aligned” with standards such as those of the Iowa Core?

A: If there are no limits on how long the test is or how much instructional time is used for testing, then full alignment is possible. Test developers balance the goal of full alignment with many other factors important to educators. The degree of alignment depends on the balance required for the intended purpose of testing.

Q: What is the Iowa Core?

A: The Iowa Core is the set of education standards adopted by the State Board of Education in July 2010. In English Language Arts and Mathematics, it consists of the Common Core State Standards, which were the standards required by the U.S. Department of Education for states to be eligible to compete for Race to the Top funds. In the areas of ELA and mathematics, the Iowa Core 2010 replaced the previous version that was adopted in 2008.

Q: What is covered by the Iowa Core?

A: The Iowa Core contains specific standards at each grade level. In English Language Arts, there are 19 standards per grade level. In Mathematics, there are between 24 and 68 standards per grade level, for a total of 272 standards.

Q: What purpose has been served by the *Iowa Assessments* for the state of Iowa.

A: Since 2003, the *Iowa Assessments* were used to provide the necessary reporting components for the federally mandated NCLB in Reading, Mathematics and Science for students in grades 3-8 and 11. Prior to NCLB, the *Iowa Assessments* were used by schools in the state to evaluate achievement and monitor growth of students. Current Iowa Code mandates assessment in Reading and Mathematics in grade 4, 8, and 11 and in science in grades 8 and 11.

Q. Was ITP aware of the Department's plan to align Form E to the 2010 Iowa Core?

A: As part of Iowa Testing Program's (ITP) contract with the Iowa Department of Education to deliver, score and report its state testing program, ITP reviewed the RFP that initiated this project and provided feedback to the Department concerning the inappropriateness of the selected form at that time.

Q: Was ITP aware of the information in the report prior to its release?

A: As part of Iowa Testing Program's (ITP) contract with the Iowa Department of Education to deliver, score and report its state testing program, ITP was charged with preparing the test materials, calculating item-level statistics, and reviewing the results for integrity and soundness. ITP's concerns about the integrity of the report were shared with the DE in June, 2013.

Timelines

Q: Is an alignment study specific to a test form (a collection of test questions)?

A: Yes, alignment studies match specific sets of test questions (a form) to a set of standards.

Q: Why was Form E selected for the alignment study?

A: It is not clear why Form E was selected for this study. Study results (page 5 of the Overview and page 15 of the Mathematics results) publically acknowledge that the appropriate standards were not used for the alignment study. Excerpts from these two pages of the report are provided below.

The purpose of each alignment study was to determine the degree of alignment between the Iowa Core Standards for each grade and the test items found on the corresponding grade-level Iowa Assessment. ***It should be noted that the Iowa Core Standards used for the alignment process were not the standards used to develop the Iowa Assessments.*** *The Iowa Department of Education provided the Iowa Core Standards to be used for the study. Page 5*

Reviewers were asked to align the assessment items to the Iowa Common Core. As previously mentioned in the Overview, ***the standards of the Iowa Common Core were not the standards used to develop the test but were used in this study as requested by the Iowa Department of Education. Using standards that were not used to develop the items could affect the results of the alignment criteria.*** *Page 15*

Q: Why was Form E introduced in the state of Iowa?

A: Form E was built to the Iowa Core as the standards were mandated in 2008. This version of the Iowa Core was based on grade spans (primary, intermediate, middle and high school). Form E of the *Iowa Assessments* was built to transition between the adoption of the Iowa Core in 2008 and the full implementation of the Iowa Core in 2015, while maintaining the reporting requirements of NCLB.

Q: When was the Iowa Core modified to include the common core state standards?

A: It was adopted in July of 2010, and fully implemented by 2015.

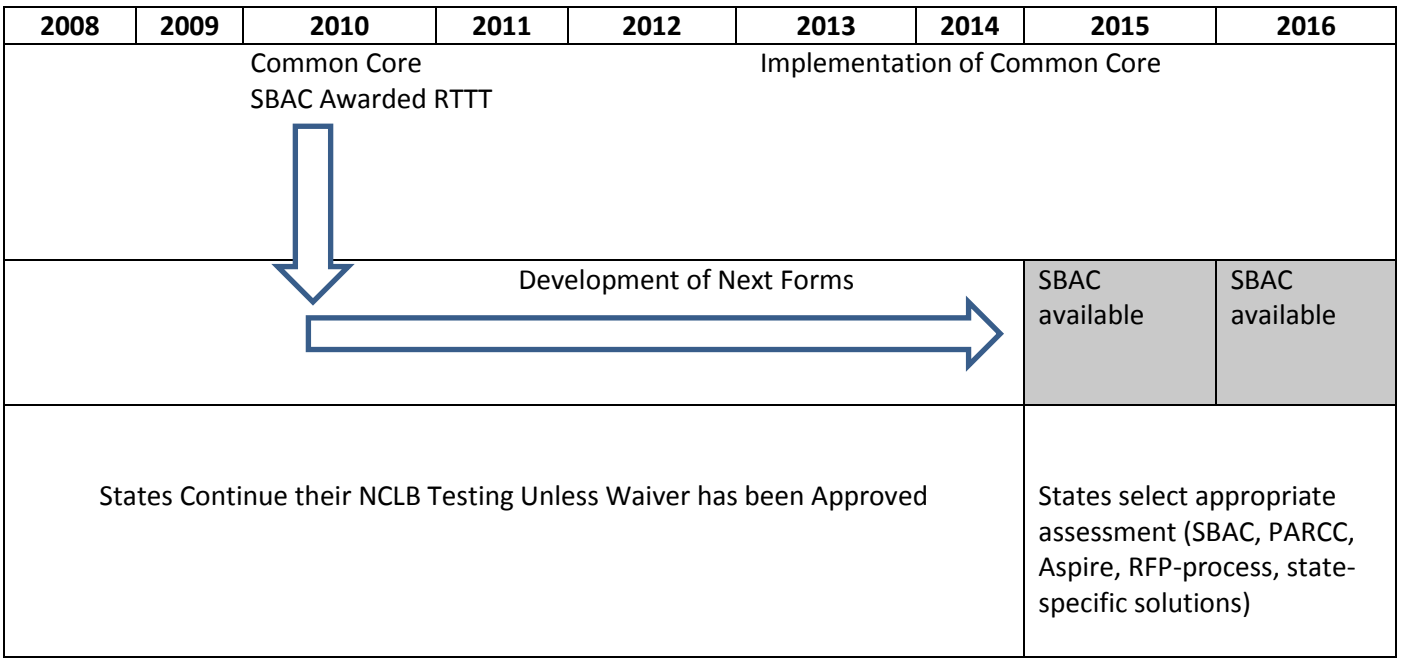
Other Assessments—Smarter Balanced Assessment Consortium (SBAC)

Q: When will SBAC be available to measure the 2010 version of the Iowa Core?

A: In the spring of 2015.

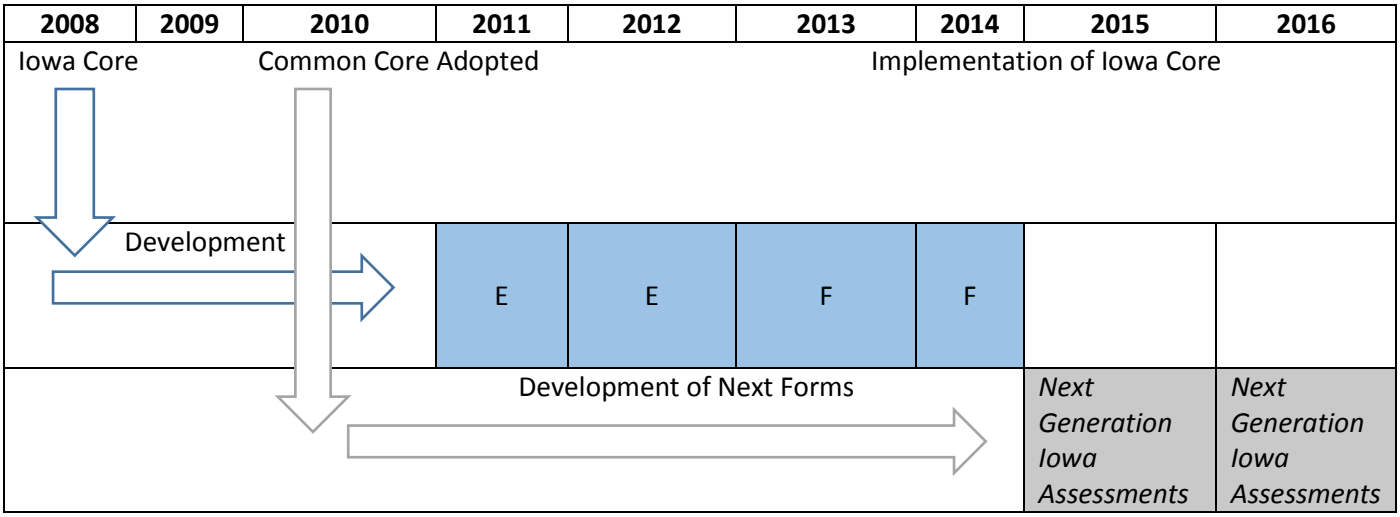
Q: Why isn't SBAC available to measure the 2010 version prior to 2015?

A: Test development, field testing, validation studies and other research require several years to design and implement. See figure below.



Q: When will the *Next Generation Iowa Assessments* be available to measure the 2010 version of the Iowa Core?

A: In the spring of 2015. The graphic below shows the interaction of the Iowa Core (2008), the Iowa Core (2010) and the introduction of new *Iowa Assessments*.



Q: Will the *Next Generation Iowa Assessments* be more fully aligned to the Iowa Core?

A: The *Next Generation Iowa Assessments*, planned for an introduction in 2015, will include additional item types (extended performance assessments, short constructed-response items and technology-enhanced items) to provide a more comprehensive alignment.

The *Next Generation Iowa Assessments* will not include any test questions that span different grades; rather all questions will be grade-level specific.

Use of Alignment Study Results

Q: Is the methodology of this study sound?

A: Judges in this study achieved such poor levels of agreement that the only trustworthy conclusion supported by the data is that the training was inadequate to the task. The poor reliability of the judges' ratings, acknowledged by its authors (Tables 14 and 15 of the report), precludes any absolute statements about alignment of the *Iowa Assessments*, Form E. These reliability results range from .17 to .46 in mathematics on a scale from 0 to 1, where 1 would indicate all the readers completely agreed with one another. Acceptable reliabilities for achievement tests are usually above .88. The study's authors themselves state that agreement levels below .5 are "poor."

Q: Should the results of this study be used to make a recommendation concerning future assessments in the state of Iowa?

A: Prior to the implementation of a new assessment in the spring of 2017 (Senate File 215, April 2013), the state of Iowa will need to compare the alignment, reliability, validity, growth indicators, readiness indicators and cost of the SBAC, the *Next Generation Iowa Assessments* and all other assessments under consideration by the state. The current study is not relevant to that purpose.

Q: Is it appropriate to use the results of the alignment study to make a recommendation about SBAC?

A: No, an alignment study that compares the SBAC to the Iowa Core will need to be completed prior to making a recommendation.

Q: Will the SBAC assessments be aligned using this same methodology?

A: SBAC will need to address this. However at this time, no SBAC tests exist that can be subjected to an alignment study.

Alignment Study Results

Q: Why does the alignment study suggest that there is weak or no alignment between the *Iowa Assessments* in Mathematics and the Iowa Core in the Categorical Concurrence area?

A: There are two responses to this question. First, the reported results were unacceptably inconsistent and low and would question the integrity of the conclusion. The figure below captures the reliability estimates contained in Tables 14 and 15 of the report. As indicated in the text of the report, these results (all less than .5) indicate poor agreement between the panelists.

Webb's methodology uses pairwise agreement to calculate the agreement of how the reviewers assigned the standard to the items. The pairwise agreement is averaged over all the assessment items to give the pairwise agreement statistic for the grade as a whole. A pairwise comparison result of 0.7 or higher reflects good agreement, a result of 0.6 or higher is considered to reflect a reasonable agreement, and a result of less than 0.5 is considered to reflect poor agreement (Webb, 2005).

The pairwise agreements for mathematics and reading are below. The reading is updated using the adjusted judgments as discussed with Iowa Department of Education.

Table 14: Summary of Pairwise—Mathematics

Grade	Number of Reviewers	Pairwise Agreement
3	7	0.457341
4	7	0.362084
5	7	0.327381
6	7	0.289683
7	8	0.252296
8	8	0.221905
10	8	0.170536
11	8	0.174107

Table 15: Summary of Pairwise—Reading

Grade	Number of Reviewers	Pairwise Agreement
3	7	0.524971
4	7	0.369615
5	7	0.297896
6	7	0.275433
7	6	0.478519
8	6	0.498551
9-10	6	0.425
11-12	6	0.380833

Second, a transitional form of the *Iowa Assessments* was included in this study. Form E was built to the 2008 Iowa Core. Table 5 of the alignment report indicates that the Mathematics alignment study attempted to match test questions to 272 math standards. This was a significant increase in the number of standards since the Iowa Core was adopted in 2008, again raising serious doubts as to the veracity of the results of the alignment study.

Grade	Number of Standards in Iowa Core (2008)	Number of Standards in Iowa Core (2010)
3	23	25
4		28
5		26
6	20	29
7		24
8		28
10	17	68
11		44

Q: Why does the alignment study suggest that there is weak or no alignment between the *Iowa Assessments* in Mathematics and the Iowa Core in the area of Balance of Representation.

A: Balance of Representation (as defined by the study) emphasizes the relative emphasis of the standards on the test and should reflect the emphasis of the state's curriculum. The study considered all standards to be of equal importance. So, for a 10th grade student, this study suggests that all 66 standards are equal in importance.

Q: Why were there significant differences between the reported alignment of the *Iowa Assessment* in Reading and Mathematics?

A: Table 10 of the report indicates that the Reading alignment was an attempt to match test questions to 152 standards, 19 at each grade level. These 19 were more consistent with the 2008 Iowa Core reading standards. There is also much greater overlap between grade levels in the reading standards than there is in the Mathematics standards.

Grade	Number of Standards in Iowa Core (2008)	Number of Standards in Iowa Core (2010)
3	7	19
4		19
5		19
6	6	19
7		19
8		19
10	6	19
11		19

Q: Were the alignment panelists consistently able to align test questions to standards in Mathematics?

A: No. The consistency and reliability of the raters was unacceptably low. At times, none of the raters could agree on the appropriate domain to place the item into. At other times, none of the raters could agree on the grade level in which to place the item.

> Steve, Adam and Angelica split a $10.84 pizza bill evenly. Which expression represents how much each person paid in dollars?
>
> A. 10.84 – 2
> B. 10.84 – 3
> C. 10.84 ÷ 2
> D. 10.84 ÷ 3

For example, as part of this study, 15 independent raters (seven from one panel and eight from a second panel), all trained in the Iowa Core, were asked to align the same items (Table D2 of the report) to the Iowa Core. The question presented in the sidebar was aligned by the 15 raters to 11 different standards, cutting across four grade levels and four content domains. The various standards that were identified are listed in the table below, along with the number of raters that identified each standard.

The item was classified as an Operations and Algebraic Thinking item by Iowa Testing Programs.

Grade	Domain	Standard	Raters
4	Measurement and data	Use the four operations to solve word problems involving distances, intervals of time, liquid volumes, masses of objects, and money, including problems involving simple fractions or decimals, and problems that require expressing measurements given in a larger unit in terms of a smaller unit. Represent measurement quantities using diagrams such as number line diagrams that feature a measurement scale.	1
5	Numbers and Operations in Base Ten	Add, subtract, multiply, and divide decimals to hundredths, using concrete models or drawings and strategies based on place value, properties of operations, and/or the relationship between addition and subtraction; relate the strategy to a written method and explain the reasoning used.	1
6	Expressions and Equations	Write and evaluate numerical expressions involving whole-number exponents.	1
7	Expressions and Equations	Solve multi-step real-life and mathematical problems posed with positive and negative rational numbers in any form (whole numbers, fractions, and decimals), using tools strategically. Apply properties of operations to calculate with numbers in any form; convert between forms as appropriate; and assess the reasonableness of answers using mental computation and estimation strategies.	1
4	Operations and Algebraic Thinking	Multiply or divide to solve word problems involving multiplicative comparison, e.g., by using drawings and equations with a symbol for the unknown number to represent the problem, distinguishing multiplicative comparison from additive comparison.1	1
5	Operations and Algebraic Thinking	Write simple expressions that record calculations with numbers, and interpret numerical expressions without evaluating them. For example, express the calculation “add 8 and 7, then multiply by 2” as 2 × (8 + 7). Recognize that 3 × (18932 + 921) is three times as large as 18932 + 921, without having to calculate the indicated sum or product.	5
6	Number System	Fluently add, subtract, multiply, and divide multi-digit decimals using the standard algorithm for each operation.	1
7	Number System	Solve real-world and mathematical problems involving the four operations with rational numbers.1	1
6	Ratios and Proportional Relationships	Use ratio and rate reasoning to solve real-world and mathematical problems, e.g., by reasoning about tables of equivalent ratios, tape diagrams, double number line diagrams, or equations.	1
7	Ratios and Proportional Relationships	Compute unit rates associated with ratios of fractions, including ratios of lengths, areas and other quantities measured in like or different units. For example, if a person walks 1/2 mile in each 1/4 hour, compute the unit rate as the complex fraction 1/2/1/4 miles per hour, equivalently 2 miles per hour.	1
7	Ratios and Proportional Relationships	Recognize and represent proportional relationships between quantities.	1
4 Grades	**6 Domains**	**Totals**	**15**

Q: Why were some of math questions included in the *Iowa Assessments* reported to be aligned to multiple grades?

A: In 2008, the mandated Iowa Core included standards that were clustered by grade spans (primary elementary, intermediate elementary, middle school and high school).

Q: How would variability in time-of-year testing affect the results of the alignment study.

A: At the item level, time-of-year testing would not affect the alignment. However, it is worth noting that the state of Iowa workbook filed for NCLB in 2002, argued that flexibility in time-of-year testing should be maintained in the state of Iowa. This flexibility placed additional demands on the items selected for Forms E and F to maintain the reporting scale, measure growth across years, and also allow flexible time-of-year testing.

Q: Would you expect the *Iowa Assessments* to be aligned with mathematics standards the Iowa Core?

A: Form E was designed to measure the Iowa Core (legislatively mandated in 2008). The four essential mathematics strands of the Iowa Core (2008), Number and Operations, Algebra, Geometry and Measurement, and Data Analysis and Probability, and the specified grade spans were used.

Given that the most significant difference between the Iowa Core (2008) and the Common Core (2010) was the placement of content to specific grade levels, we expect that all items found on Form E align with the content of the Common Core, but that the grade level would vary. This expectation is supported by the results in Table 6 of the report that states only 5 of the 272 math items did not align to a new math standard.