Utah Aspire Plus
2019–2020 Technical Bulletin

TABLE OF CONTENTS

List of Tables

1. Introduction

1.1 Background

The Utah Aspire Plus summative assessments were created out of Utah Statute 53E-4-304 (https://le.utah.gov/xcode/Title53E/Chapter4/53E-4-S304.html?v=C53E-4-S304_2019051420190514). The statute requires the Utah State Board of Education (USBE) to administer assessments that are predictive of college readiness at grades 9 and 10 and provide a growth score from grade 9 to 10. The Utah Aspire Plus assessments are a hybrid of ACT Aspire and Utah Core test items. These are computer-based, fixed-length tests intended to measure end-of-grade-level high school knowledge and skills for students in grades 9 and 10. Spring 2019 marked the first administration of the Utah Aspire Plus assessments and the creation of base reporting scales for each respective grade and subject assessment.

Prior to 2019, students were assessed on the core standards through the Utah Student Assessment of Growth and Excellence (SAGE) assessment program. The Utah Aspire Plus assessment program is an extension of the Utah SAGE, still intended to measure student performance in relation to the Utah Core Standards (https://www.uen.org/core/), but also intending to measure students' preparedness for meeting college readiness benchmarks. As such, the assessment content from Utah SAGE is used as one component of the Utah Aspire Plus assessments.

Additional content from ACT Aspire is used to provide predictions of performance on the ACT®. The ACT® is the primary college readiness assessment submitted to local universities in Utah. All juniors in Utah public schools take the ACT. The ACT Aspire content also aligns to the Utah Core Standards and is counted toward Utah Aspire Plus scores. As such, the Utah Aspire Plus assessments incorporate test questions from the ACT Aspire assessments that are used not only to contribute to student overall scores but also to provide a predictive indicator of performance on the ACT®. Students receive predicted ACT® score ranges for each ACT® subtest (English, reading, mathematics, and science), as well as an overall predicted composite ACT® score range.

As required by the statute noted previously, the assessments also provide overall scores as indicators of end-of-grade-level expectations for 9th and 10th grade students and performance level indicators (*Below Proficient*, *Approaching Proficient*, *Proficient*, and *Highly Proficient*) for English, reading, mathematics, and science.

As stated, the first operational administration was conducted in the spring of 2019 at grades 9 and 10 for English, reading, mathematics, and science. Data from the inaugural administration were used to establish the initial Utah Aspire reporting scales and the setting of performance levels. Technical details of these features and activities are presented in the *2018-2019 Utah Aspire Plus Technical Report* (http://utah.pearsonaccessnext.com/resources/additional-services/2018-19%20UA+%20Tech%20Report_Web.pdf). Spring 2020 was intended to mark the second operational administration of the Utah Aspire Plus tests. In spring of 2020, Senate Bill 3005, which included a waiver of the Utah Aspire Plus assessment requirements, was passed during the Utah Legislature's 3rd Special Session of 2020 and signed into law on April 22, 2020. As a result, the spring testing of Utah Aspire Plus was cancelled. This technical bulletin is intended to describe relevant activities that still occurred for the Utah Aspire Plus program.

1.2 Purpose of the Operational Tests

The Utah Aspire Plus assessments are designed for several purposes. First, the tests are intended to measure the breadth and depth of the Utah Core Standards and measure across all levels of student performance. Second, the tests are created to provide awareness of individual achievement in relation to stated performance expectations. Third, performance on the tests is intended to provide evidence of whether students are on track for college and career readiness. Finally, the tests are used to evaluate growth between 9th and 10th grade.

1.3 Composition of the 2020 Operational Tests

Each operational Utah Aspire Plus test form was constructed to reflect the full test blueprint in terms of content, standards measured, and item types. All blueprints were designed to measure knowledge and skills described in the Utah Core Standards (https://www.uen.org/core/). For science, the Utah Aspire Plus blueprints are further explicated to measure 1) science content specific to biology, chemistry, Earth science, or physics; and 2) Intended Learning Outcomes (ILOs). The ILOs describe the goals for science skills and attitudes. They are defined for each grade and are an integral part of the standards that are used to guide science instruction (https://www.schools.utah.gov/File/8cf206d1-022d-42ec-b02d-3cbad59ecb79). Additionally, the tests are designed to focus on the underlying skills of science as defined in the ILOs (e.g., science process and thinking skills, etc.) and not require specific knowledge of the scientific discipline (meaning a chemistry student ought to possess the skills necessary to answer a biology question).

The Utah Aspire Plus tests are composed of several different types of items that are intended to elicit evidence of a student's knowledge and skills in a variety of ways. These include multiple choice, multiple select, evidence-based selected response, and technology enhanced (TE). Multiple-choice items present students with four or five responses, of which there is one correct answer. Multiple-select items require students to select two or three correct choices from several presented choices. These items combine a traditional selected-response question (Part A) with a second selected-response question (Part B) that asks students to show evidence from the text that supports the answer they provided to the first question. In this manner, Part A is designed as an *identification* component, where Part B is designed to elicit an *evidence*-based component. Further, these types can be designed as two multiple-choice items, or a combination of multiple-choice and technology-enhanced (TE) items. Technology-enhanced (TE) items require specialized interactions within the online presentation for capturing student responses (e.g., drag and drop).

The Utah Aspire Plus English tests target language conventions and comprehension. Students should be able to demonstrate command of standard English grammar, usage, capitalization, punctuation, and spelling. In addition, students should be able to demonstrate vocabulary knowledge in comprehending complex texts.

The Utah Core Standards in Reading define expectations of comprehension skills, understanding tone and point of view of texts, and evaluating texts. On the Utah Aspire Plus Reading tests, students must demonstrate these skills with different types of text sources.

The assessment context for Utah Aspire Plus Mathematics is grounded in five conceptual categories from the Utah Core Standards: Number and Quantity, Algebra, Functions, Geometry, and Statistics and Probability. There are two general levels of math content for Utah Aspire Plus. The first level, referred to as Secondary Math I, extends the mathematics from the middle school grades, particularly on linear and exponential relationships. The next level, Secondary Math II, focuses on quadratic relationships and comparing them to the linear and exponential relationships from Secondary Math I.

As noted above, the primary emphasis of the Utah Aspire Plus Science tests is on the Intended Learning Outcomes (ILOs), which describe the skills students should learn from science instruction. From ILOs, students should use science as a process of obtaining knowledge based upon observable evidence. As noted, these skills are applicable regardless of domain (i.e., Biology, Physics, etc.).

1.4 Intended Population of the Operational Tests

The Utah Aspire Plus tests are designed for students completing their 9th and 10th grade courses in English Language Arts (ELA), mathematics, and science. The English and reading tests are designed to assess the skills that 9th and 10th grade ELA students should have by the end of those respective years. The mathematics tests are designed to assess the skills that 9th (Secondary Math I) and 10th grade (Secondary Math II) math students should have by the end of those respective years. The science tests are designed to assess the skills that 9th and 10th grade students taking biology, chemistry, Earth science, or physics should have by the end of instruction (regardless of the specific course).

1.5 Overview of the Technical Bulletin

The intended audience of the report are those with a basic technical understanding of large-scale assessment systems and their uses. It assumes some technical knowledge of how score scales are developed and derived and how scores are intended to support valid interpretations of intended claims.

This bulletin provides details of activities involved in creation of the second year of the Utah Aspire Plus testing system at grades 9 and 10. While the report does not include technical details regarding outcomes of the test administration, it does include additional details regarding further development of content related to a spring 2021 operational administration (including the new science assessments).

2. Test Development

2.1 Overview of the 2020 Utah Aspire Plus Assessments, Claims, and Blueprints

The Utah Aspire Plus assessments are aligned to the Utah Core Standards and designed to measure the breadth and depth of the Utah Core Standards across all levels of student performance, to provide awareness of individual achievement in relation to stated performance expectations, and to provide evidence of whether students are on track for college and career readiness. Utah Aspire Plus content follows a rigorous development process that meets and often

exceeds industry standards for best practices in assessment. Every item, written by Utah teachers, goes through an extensive review designed to ensure adherence to high quality and the principles of universal design.

This chapter describes the claims intended to support the purposes outlined in Chapter 1; the development of blueprints defining the components of the Utah Aspire Plus assessments that reflect the breadth of the Utah Core Standards across different levels of student understanding; and the development of tasks (items) intended to fulfill the respective blueprints and provide evidence of varying levels of performance reflective of each of the stated claims.

It should be noted that while both claims and sub claims are presented here for each subject, only the claims are reported on individual student reports (ISR). Sub claims currently only provide structure within the respective blueprints but are <u>not</u> reported at the individual student level.

2.1.1 English Assessment Claims

The Utah Aspire Plus English tests target language conventions and comprehension. Students should be able to demonstrate command of standard English grammar, usage, capitalization, punctuation, and spelling. In addition, students should be able to demonstrate vocabulary knowledge in comprehending complex texts.

The claim structure for the Utah Aspire Plus English tests is drawn from the Utah Core Standards and frames the design and development of the summative tests at grades 9 and 10.

> **Claims:** The primary claims reflect the main goals for the use of the Utah Aspire Plus English tests. The first is that student performance reflects an indicator of career and college readiness as demonstrated through students' understanding of language conventions and comprehension as expected to have been attained by the end of each respective year as a prediction of performance on the ACT® English test. Second is that overall performance reflects students' understanding of language conventions and comprehension with respect to the breadth and depth of the Utah Core Standards and measures across all levels of student performance.
>
> **Sub Claims:*** The sub claims further explicate what is measured on Utah Aspire Plus English tests and are grouped into the following categories:
>
> - Production of Writing
> - Knowledge of Language
> - Conventions of Standard English

* It should be noted that sub claims are *not* reported on individual student reports but form an important structural element within the blueprints. They are included in this technical report for completeness.

2.1.2 Reading Assessment Claims

The Utah Aspire Plus Reading tests define expectations of comprehension skills, understanding tone and point of view of texts, and evaluating texts. On the Utah Aspire Plus Reading tests, students must demonstrate these skills with different types of text sources.

The claim structure for the Utah Aspire Plus Reading tests is drawn from the Utah Core Standards and frames the design and development of the summative tests at grades 9 and 10.

> **Claims:** The primary claims reflect the main goals for the use of the Utah Aspire Plus Reading tests. The first is that student performance reflects an indicator of career and college readiness as demonstrated through students' ability to read and comprehending complex informational and literary texts as expected to have been attained by the end of each respective year as a prediction of performance on the ACT® Reading test. Second is that overall performance reflects students' understanding of reading and comprehending complex informational and literary texts with respect to the breadth and depth of the Utah Core Standards and measures across all levels of student performance.
>
> **Sub Claims:*** The sub claims further explicate what is measured on Utah Aspire Plus Reading tests and are grouped into the following categories:
>
> - Key Ideas
> - Craft and Structure
> - Integration of Knowledge and Ideas

2.1.3 Mathematics Assessment Claims

The Utah Aspire Plus Mathematics tests are grounded in five conceptual categories from the Utah Core Standards: Number and Quantity, Algebra, Functions, Geometry, and Statistics and Probability. There are two levels of math content for Utah Aspire Plus that reflect expectations at grades 9 and 10, respectively. The first level (grade 9), referred to as Secondary Math I, extends the mathematics from the middle school grades, particularly on linear and exponential relationships. The next level, Secondary Math II (grade 10), focuses on quadratic relationships and comparing them to the linear and exponential relationships from Secondary Math I.

The claim structure for the Utah Aspire Plus Mathematics tests is drawn from the Utah Core Standards and frames the design and development of the summative tests at grades 9 and 10.

> **Claims:** The primary claims reflect the main goals for the use of the Utah Aspire Plus Mathematics tests. The first is that student performance reflects an indicator of career and college readiness as demonstrated through students' ability to understand linear relationships, abstract and quantitative reasoning, and problem solving as expected to have been attained by the end of each respective year as a prediction of performance on the ACT® Mathematics test. Second is that overall performance reflects students' understanding of linear relationships, abstract and quantitative reasoning, and problem solving with respect to the breadth and depth of the Utah Core Standards and measures across all levels of student performance.

Sub Claims:* The sub claims further explicate what is measured on Utah Aspire Plus Math tests and are grouped into the following categories:

Math I (Grade 9)

- Algebra
- Functions
- Geometry
- Statistics and Probability

Math II (Grade 10)

- Number and Quantity
- Algebra
- Functions
- Geometry
- Statistics and Probability

2.1.4 Science Assessment Claims

The Utah Aspire Plus Science tests are developed around the Utah Core Standards for science as described in the Intended Learning Outcomes (ILOs). From ILOs, students are expected to use science as a process of obtaining knowledge based upon observable evidence. As noted, these skills are applicable regardless of domain (Biology, Physics, Earth Science, and Chemistry).

The claim structure for the Utah Aspire Plus Science tests is drawn from the Utah Core Standards as described in the ILOs and frames the design and development of the summative tests at grades 9 and 10.

Claims: The primary claims reflect the main goals for the use of the Utah Aspire Plus Science tests. The first is that student performance reflects an indicator of career and college readiness as demonstrated through students' ability to understand and apply science as defined by the ILOs as expected to have been attained by the end of each respective year as a prediction of performance on the ACT® Science test. Second is that overall performance reflects students' understanding of science as defined by the ILOs with respect to the breadth and depth of the Utah Core Standards and measures across all levels of student performance.

Sub Claims:* The sub claims further explicate what is measured on Utah Aspire Plus Science tests and are grouped into the following categories:

- ILO 1 – Use Science Process and Thinking Skills
- ILO 3 – Demonstrate Understanding of Science Concepts, Principles, and Systems
- ILO 4 – Communicate Effectively Using Science Language and Reasoning
- ILO 5/6 – Demonstrate Awareness of Social and Historical Aspects of Science/Demonstrate Understanding of the Nature of Science

2.2 Utah Aspire Plus Blueprint Creation

The Utah Aspire Plus tests are administered in English, reading, mathematics, and science in grades 9 and 10 and are described in Section 1.3. For the Utah Aspire Plus tests, the creation of test blueprints was driven by the intended purposes detailed previously in order to support the respective claim structures. The blueprints for Utah Aspire Plus are the distribution of item types across domains/reporting categories, level of cognitive demand, and the number of total points associated with each. Details of the creation of the Utah Aspire Plus blueprints are provided in the *2018-2019 Utah Aspire Plus Technical Report*.

2.3 Test Development Activities

Prior to the creation of Utah Aspire Plus, students were tested on the Utah Core Standards through the Utah Student Assessment of Growth and Excellence (SAGE). The Utah Aspire Plus assessments were built from existing Utah SAGE banked content combined with items from ACT Aspire to allow for predictions of students' preparedness for meeting college readiness. All available content for creation of the 2020 Utah Aspire Plus tests was based on the existing item banks described in the *2018-2019 Utah Aspire Plus Technical Report*.

For creation of the 2020 tests, two important design elements are worth noting. The first is that sets of items administered in 2019 were selected to serve as *linking* or *common* items that would be used to equate the 2020 Utah Aspire Plus tests to the 2019 base scales within a common item non-equivalent groups equating design (Kolen and Brennan, 2014). For test development purposes, this meant selecting sets of items to ideally reflect a miniature version of the overall test (typically at least 20 percent) in content as well as statistical characteristics. The second element worth noting is that a different set of ACT Aspire content was used for this second-year forms development activity. This helped limit exposure of the Aspire content that might otherwise negatively impact ACT predication score activities. However, it also meant linking sets used for equating did not have any ACT content available to serve as common items for the 2020 test forms. Still, final linking sets that reflected at least 20 percent of the overall tests and of comparable content were able to be selected for the Utah Aspire Plus tests.

2.3.1 Operational Forms Development

The construction of test forms for the 2020 Utah Aspire Plus was a coordinated effort between experts from the Utah State Board of Education, Pearson, and ACT. This process required adhering to guidelines that promote fair and ethical testing practices. Using the content developed to measure the Utah Core Standards, specialists worked through an iterative process to evaluate the specific items, passages, and stimuli that best met the intended measurement targets and to support all stated claims.

The Utah Aspire Plus assessments measure students' mastery of the Utah Core Standards and science ILOs. These standards are used to drive Utah instruction as well as developing the Utah Aspire Plus tests. As stated earlier, the Utah Aspire Plus assessments are designed so that test scores can be linked to ACT scales to provide students with indicators of being prepared for meeting college readiness benchmark. In order to accomplish this, approximately 50% of the Utah Aspire Plus tests are composed of items from ACT Aspire. As noted, these items serve

multiple purposes, which include being used to derive prediction scores between the Utah Aspire Plus scales and ACT scales.

The general test development process for Utah Aspire Plus was initiated with the selection of items from ACT Aspire. Items were selected based on match to blueprint, as well as statistical indicators of item quality and fairness provided from the SAGE and ACT Aspire banks, respectively. ACT Aspire items were positioned within each form in the same locations as originally administered within ACT Aspire forms to help facilitate the derivation of the predictive scores on Utah Aspire Plus.

Once the ACT Aspire items were selected, Pearson psychometrics selected sets of items common to 2019 that would be used to equate the 2020 tests to the 2019 base scales. In addition to selecting items to be as similar as possible to the overall blueprints, but they were also targeted to the original base scale difficulties.

This procedure was an iterative process whereby the first proposed form is evaluated by each party (Pearson, USBE, and ACT) for content and psychometric quality, feedback provided, and revisions made until a best final version was approved by all. It should be noted that without new development of content, bank limitations meant an inability to strictly meet the new blueprint in all cases (see below). It also meant that there were also instances where items with poorer statistical indices were included to meet the blueprint. These were infrequent and, in all cases, deemed reasonable in supporting the intended claims without negative impact. Moving forward, newly developed content will fill gaps and address such limitations as the assessments mature.

2.3.2 Statistical Guidelines

While the initial Utah Aspire Plus tests were primarily driven by content considerations, statistical indices were available based on use within the SAGE and ACT Aspire Plus assessments. For creation of Utah Aspire Plus tests, some general guidelines were used to help support selection of a range of item difficulties and evaluate item quality to ensure the best overall test forms.

The guidelines for creation of the Utah Aspire Plus forms were as follows:

- **Target item difficulty range of between 0.30 and 0.85.** Based on *p*-values, where the percentage reflects the percentage of students correctly responding to the item. Items awarding more than one point used the item mean divided by the maximum points possible to place on the *p*-value metric (average of the maximum points possible on a given item).
- **Target threshold for item discrimination of 0.20 and above.** Where item discrimination is defined by item-total score correlations.
- **Extreme differential item functioning (DIF) indices should be avoided**. A standard flagging convention indicates differences of magnitude and classifies the most extreme cases of DIF as “C,” moderate DIF as “B,” and minor to no DIF as “A.” As such, items flagged “C” should be avoided and minimal use of items flagged “B” should be used and/or balanced within a form where possible.

2.3.3 2020 Match to Test Blueprint

Tables 1 through 8 present the match between the final 2020 operational forms of Utah Aspire Plus and the test blueprints. English, reading, math, and science final forms reasonably matched all targets by item type, depth of knowledge, and reporting category (within 3 percent).

Table 1. Utah Aspire Plus English Grade 9 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	24–31	48%	62%	58%
Technology Enhanced	20–26	40%	52%	42%
Depth of Knowledge				
Level 1	22–33	44%	66%	57%
Level 2	5–12	10%	24%	16%
Level 3	12–17	24%	34%	27%
Reporting Categories				
Production of Writing	9–14	18%	28%	20%
Knowledge of Language	4–10	8%	20%	9%
Conventions of Standard English	28–38	56%	76%	71%

Table 2. Utah Aspire Plus English Grade 10 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	24–31	48%	62%	56%
Technology Enhanced	20–26	40%	52%	44%
Depth of Knowledge				
Level 1	22–33	44%	66%	54%
Level 2	5–12	10%	24%	15%
Level 3	12–17	24%	34%	30%
Reporting Categories				
Production of Writing	9–14	18%	28%	24%
Knowledge of Language	4–10	8%	20%	13%
Conventions of Standard English	28–38	56%	76%	63%

Table 3. Utah Aspire Plus Reading Grade 9 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	22–29	62%	82%	69%
Technology Enhanced	2–5	6%	14%	17%
Evidence-Based Selected Response	4–6	10%	17%	14%
Depth of Knowledge				
Level 1	4–10	11%	28%	11%
Level 2	12–20	34%	57%	49%
Level 3	9–14	25%	40%	40%
Reporting Categories				
Key Ideas	9–18	26%	51%	51%
Craft and Structure	14–20	40%	57%	37%
Integration of Knowledge and Ideas	3–5	9%	14%	11%

Table 4. Utah Aspire Plus Reading Grade 10 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	22–29	62%	82%	83%
Technology Enhanced	2–5	6%	14%	5%
Evidence-Based Selected Response	4–6	10%	17%	11%
Depth of Knowledge				
Level 1	4–10	11%	28%	14%
Level 2	12–20	34%	57%	47%
Level 3	9–14	25%	40%	39%
Reporting Categories				
Key Ideas	9–18	26%	51%	50%
Craft and Structure	14–20	40%	57%	38%
Integration of Knowledge and Ideas	3–5	9%	14%	11%

Table 5. Utah Aspire Plus Mathematics Grade 9 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	30–33	75%	83%	75%
Technology Enhanced	7–10	18%	25%	25%
Depth of Knowledge				
Level 1	8–12	20%	30%	28%
Level 2	15–20	38%	50%	50%
Level 3	9–13	23%	33%	23%
Reporting Categories				
Algebra	9–11	23%	28%	28%
Functions	10–12	25%	30%	28%
Geometry	9–11	23%	28%	25%
Statistics and Probability	7–9	18%	23%	20%

Table 6. Utah Aspire Plus Mathematics Grade 10 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	30–33	75%	83%	78%
Technology Enhanced	7–10	18%	25%	23%
Depth of Knowledge				
Level 1	8–12	20%	30%	30%
Level 2	15–20	38%	50%	48%
Level 3	9–13	23%	33%	23%
Reporting Categories				
Number and Quantity	2–4	5%	10%	10%
Algebra	9–11	23%	28%	25%
Functions	10–12	25%	30%	28%
Geometry	11–13	28%	33%	30%
Statistics and Probability	2–4	5%	10%	8%

Table 7. Utah Aspire Plus Science Grade 9 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	29–34	81%	94%	93%
Technology Enhanced	2–3	6%	8%	8%
Depth of Knowledge				
Level 1	3–9	8%	25%	18%
Level 2	12–23	33%	64%	53%
Level 3	8–13	22%	36%	30%
Reporting Categories				
(ILO) 1: Use Science Process and Thinking Skills	15–23	42%	64%	60%
(ILO) 3: Demonstrate Understanding of Science Concepts, Principles, and Systems	4–6	11%	17%	10%
(ILO) 4: Communicate Effectively Using Science Language and Reasoning	7–10	19%	28%	20%
(ILO) 5/6: Demonstrate Awareness of Social and Historical Aspects of Science/Demonstrate Understanding of the Nature of Science	3–4	8%	11%	10%

Table 8. Utah Aspire Plus Science Grade 10 Operational Test Blueprint Match

	Number of Items	Minimum %	Maximum %	2020 Form
Item Type				
Multiple Choice	29–34	81%	94%	93%
Technology Enhanced	2–3	6%	8%	8%
Depth of Knowledge				
Level 1	3–9	8%	25%	23%
Level 2	12–23	33%	64%	58%
Level 3	8–13	22%	36%	20%
Reporting Categories				
(ILO) 1: Use Science Process and Thinking Skills	15–23	42%	64%	55%
(ILO) 3: Demonstrate Understanding of Science Concepts, Principles, and Systems	4–6	11%	17%	13%
(ILO) 4: Communicate Effectively Using Science Language and Reasoning	7–10	19%	28%	23%
(ILO) 5/6: Demonstrate Awareness of Social and Historical Aspects of Science/Demonstrate Understanding of the Nature of Science	3–4	8%	11%	10%

2.4 Spring 2021 Utah Aspire Plus Administration

After testing was cancelled in spring 2020, it was decided that the Utah Aspire Plus test forms would be rolled over for use in the spring 2021 test administration cycle in English, reading, and math. Additionally, items were developed over the course of the 2020 testing cycle along the originally planned schedule and will be field tested as part of the spring 2021 administration.

In preparation for the 2021 field test, the Pearson Content team performed a gap analysis using metadata and item statistics from the Utah Aspire Plus item banks. Once gaps were identified each content lead developed an Item Development Plan (IDP). Each IDP considered Utah's program requirements, scope, and needs for that particular content area and grade. Materials and training documents were created and updated for use in the item writer workshop involving Utah educators.

The IDPs informed the item writer assignments for the workshop, making for a very targeted, deliberate development process. Utah educator content and fairness committees reviewed all developed items and made editorial recommendations. USBE and Pearson reconciled the feedback and USBE provided their final item approval before test construction.

2.4.1 Utah Aspire Plus Science with Engineering Education Standards Summative Assessment

The Utah Aspire Plus Science with Engineering Education Standards (SEEds) summative assessment will be administered to Utah students beginning with the spring 2021 statewide test administration. This assessment is composed of test units that assess multi-dimensional knowledge and skill interactions across different scientific phenomena within core disciplines. After test administration, performance standards will be set through a standard setting process and additional analyses will provide predicted score ranges to the ACT®.

The test units of the SEEds assessments are item clusters. Item clusters are composed of two parts: stimulus and items. The stimulus is composed of a scientific phenomenon, a context where students engage with and make sense of the phenomenon, and a task statement directing students on how to respond to the items. The items are where students engage in science practices while demonstrating understanding of disciplinary core ideas and cross cutting concepts. Three categories of item types will be used within the clusters: multiple choice and/or multiple select (MC and/or MS), evidence-based selected response, and technology enhanced items (or TEIs). Each cluster will have five or six items.

Scoring on the SEEds assessment will be done through what are called scoring assertions. For each item, these are explicit assertions made about the knowledge and skills that a student has demonstrated based on specific features of the student’s response and how demonstration of those skills relates to earning a given score point. For a multiple-choice item, the scoring assertion is based on a single cognitive task and is worth one point. A graphing item, however, may require two cognitive tasks – identification and plotting – and, thus, will be worth two points.

2.4.2 SEEds Blueprint

The SEEds blueprint assumes a design in which one of the three DCIs will be assessed by two clusters and the other two DCIs with a single cluster. Coverage of the respective DCIs will rotate across forms (either within a given year or across years) to ensure the standards are fully represented over time. For 2021 the intention is to have three forms per grade where the ACT clusters (sets of items associated with common stimuli) will serve as a common linkage across all. The following table provides the SEEds assessment test blueprint (similar at both grades).

Table 9. Utah Aspire Plus SEEds General Summative Assessment Blueprint

	Number of Items	**Number of Points/ Assertions**
Item Types		
MC, MS, 3-4 per cluster	12-16	
TEI, 2-3 per cluster	8-12	
Disciplinary Core Idea (DCI)		
Life Science	6	8-10
Earth and Space Science	12	16-20
Physical Science	6	8-10
Reporting Categories		
Gathering & Investigating (GI)	5-7	8-10
Developing Models (DM)	5-7	8-10
Using Mathematical Thinking (UM)	5-7	8-10
Construct Explanations (CE)	5-7	8-10
OP Item	**23-25**	**32-40**
FT Item	**6-8**	**8-12**
TOTAL	**29-33**	**40-52**

* Engineering standards are imbedded within specific standards.
**Field testing of clusters becomes relevant in the 2021-2022 school year.

2.5 Updated ACT Predicted Score Ranges

As noted, one of the goals of the Utah Aspire Plus assessments is to be predictive of college readiness at grades 9 and 10, and the means of this is in terms of providing prediction score ranges of performance on the ACT for the four subject tests (English, math, reading, and science) and the Composite score (the average of the four subject tests). Predicted ranges of performance were determined originally between ACT Aspire scores and ACT scores, where for a given ACT Aspire score, there was a distribution of related ACT scores. The bounds of the range were denoted by the scores closest to the 25th and 75th percentiles of the ACT score distribution, conditional on ACT Aspire scores. For Utah Aspire Plus, an additional error term was added to account for error attributable to linking the Utah Aspire Plus scores.

Students can use the predicted scores together with the ACT College Readiness Benchmarks to monitor their preparedness to be college-ready by the end of high school. Utah students take the

ACT® during their junior year of high school. Specific details from the original prediction score studies can be found in the *2018-2019 Utah Aspire Technical Report*.

In addition to relying on the relationship between the Utah Aspire Plus tests to the ACT Aspire scales for deriving the initial ACT prediction score ranges for the 2019 administration, the intention was to provide updated predictions based on longitudinal data as it becomes available. The updated ACT score ranges directly link the Utah Aspire Plus scores at grades 9 and 10 to ACT scores at grade 11. In spring 2020, the first longitudinal data was available for this purpose. The initial longitudinal Utah-to-ACT prediction studies were based on students who were in the 10th grade during the 2019 administration of the Utah Aspire Plus tests. In spring 2020 these students were in the 11th grade and most took the ACT.

Appendix A provides the details of the first longitudinal study from spring 2020. It should be noted that these results reflect preliminary findings intended for initial consideration by USBE and their technical advisory committee. Within it are described steps taken in evaluating the ACT samples in relation to previous administrations and efforts to improve predications based on a weighting procedure. Generally, these updated prediction score ranges are much tighter than the original ranges derived indirectly through the ACT Aspire scales. Similar studies are planned with longitudinal data from the spring 2021 administration, in which students who were in 9th grade in 2019 will be taking the ACT as 11th grade students.

Appendix A: Updating ACT Score Predications for Utah Grade 10 Aspire Plus

April 29, 2020 - Jeff Allen and Wei Tao, ACT

[Materials presented to USBE as part of the spring 2020 technical advisory committee meeting.]

We document the data and procedures used to generate updated ACT score predictions for the Utah Aspire Plus 10th grade assessments. Included in this documentation are:

- A description of the methodological approach
- Descriptions of the samples used to generate the predictions
- Description of weighting procedure to ensure samples are representative of 10th grade population
- Description of updated predicted ACT score ranges
    - Comparison to previously derived predicted ACT score ranges
    - Accuracy statistics

General description of methodological approach

The following steps were taken:

1) Match the spring 2019 grade 10 Utah Aspire Plus records to the spring 2020 grade 11 ACT test records. Student state ID was used to match the records.

2) Compare the matched sample to the spring 2019 grade 10 data to assess how representative the matched sample is to the target population (the spring grade 10 data is used as the target population). The matched sample is different than the target population for two primary reasons: Some students were not able to take the ACT test due to the COVID-19 school shutdowns, and some students were lost to follow-up (e.g., moved out of state, were absent on test day, or did not take the ACT for some other reason).

3) Weight the matched sample to be representative of the target population with respect to 10th grade test scores, gender, limited English proficient status, special education status, and race/ethnicity. Propensity scores, based on logistic regression models, are used to derive the weights.

4) Using the weighted data, fit the default *Student Growth Percentile* (SGP) model to estimate the percentiles of ACT scores, conditional on Utah Aspire Plus scores. The SGP model is preferred over linear regression because it does not impose assumptions of linearity or homoskedasticity.

5) Using the SGP model results, find the 25th and 75th percentiles of the ACT score distribution for each possible 10th grade test score. The 25th percentile serves as the lower bound of the predicted ACT score range and the 75th percentile serves as the upper bound of the predicted ACT score range. If ACT scores were reported on a continuous scale, this would produce predicted score ranges with 50% coverage. Because ACT scores are reported to the nearest integer, this produces predicted score ranges with approximately 60% coverage.

6) Adjust the predicted ACT score ranges to ensure that they do not decrease with Utah Aspire Plus score. This step is necessary because the SGP procedure may result in conditional percentiles that are not monotonically increasing (this only tends to happen for areas of the score distribution where the data are very sparse).

7) Assess the accuracy of the predicted ACT score ranges and compare the ranges to those that were estimated last year. Note that steps 2 through 6 are completed for each subject for which predicted ACT scores are estimated (English, math, reading, science, and composite).

Samples used to generate the predictions

Table 1 shows the number of students with reported scale scores, by subject and assessment. Note that the number of students with ACT test scores is much smaller than the number of students with 10th grade test scores because of the COVID-19 shutdown. Table 1 also provides the number of students with scale scores on both tests (Matched Sample).

Among all Utah 11th graders who took the ACT in 2020 (before the COVID-19 shutdown), the mean ACT composite score was 20.21. Among all Utah 11th graders who took the ACT in 2019, the mean ACT composite score was 19.72. The difference in mean scores is (20.21-19.72) 0.49 score points, which could be due to cohort improvement, due to differences in schools who were able to test before the shutdown, or due to both reasons. To examine this further, we compared mean ACT composite scores only for *common* schools – those schools that tested in both 2019 and 2020 and had stable N counts.[1] We found that the mean ACT composite was 19.98 in 2019 and 20.24 in 2020 for students from the common schools, suggesting cohort improvement of 0.26 score points. Thus, we believe that the overall difference in mean ACT composite scores (0.49) is partly attributable to cohort improvement, and partly attributable to schools that tested in 2020 having slightly higher mean achievement than schools that did not test.

[1] Schools were classified as having stable N counts if the number of students tested in 2019 and 2020 was within 50% of one another.

Table 1: Number of students tested and matched, by subject

Subject	Grade 10 Aspire Plus	Grade 11 ACT	Matched Sample
Composite	42,165	32,341	27,743
English	43,850	32,373	28,684
Mathematics	43,713	32,366	28,610
Reading	44,139	32,357	28,865
Science	43,917	32,348	28,708

Not all ACT-tested students had matching 10th grade records. This is expected for students who migrated into Utah after the 10th grade test, or who did not take the 10th grade test for any other reason. Among all ACT-tested students, 867 did not have a student ID, thus could not be matched. These cases of missing ID were generally spread sparsely across schools, with a few exceptions:

- High school 450030 tested over 400 students, but all students were missing student ID, thus are not included in the matched sample
- High school 450192 tested over 100 students, but 47 were missing student ID, thus are not included in the matched sample

Because the matched samples are large and representative of the target population (described later), we do not expect that this missing data would have much impact on the updated predictions.

Because the predictions are reported to 10th grade students, we used the 10th grade data as the target population. In Table 2, we compare the matched sample to the target population on 10th grade test score quintile, gender, limited English proficient status, special education status, and race/ethnicity. Table 2 only reflects the composite score analysis, but the comparison is similar across the other subject areas.

Table 2: Comparing the matched sample to the target population (composite score)

Variable	Matched Sample	Target Population	Matched Sample, Weighted
Grade 10 Aspire Plus score, quintile			
1st	16.2%	20.4%	20.5%
2nd	18.6%	19.7%	19.8%
3rd	21.1%	20.2%	20.2%
4th	22.3%	20.2%	20.2%
5th	21.7%	19.5%	19.5%
Female	49.3%	49.2%	49.0%
Limited English proficient	3.3%	4.4%	4.3%
Special education	7.6%	8.6%	8.6%
Race/ethnicity			
African American	1.1%	1.3%	1.3%
Asian	1.6%	2.0%	2.0%
Hispanic	14.2%	16.8%	16.7%
Two or more races	2.6%	2.3%	2.3%
Other	2.2%	2.6%	2.6%
White	78.3%	74.9%	75.0%

Table 2 shows that students in the matched sample tend to have higher 10th grade test scores, are slightly less likely to have limited English proficient status, are slightly less likely to have special education status, and are slightly more likely to be White. Table 2 also shows the comparison after weighting the matched sample to be more like the target population on these characteristics. After weighting, the matched sample is nearly identical to the target population. Later, we describe the method used to weight the matched sample.

Table 3 shows summary statistics for the 10th and 11th grade test scores for the matched samples and weighed matched samples. In addition to test score means and standard deviations, the correlations are also presented. The correlations range from 0.72 for reading and science to 0.86 (for composite).

Because the weighting procedure assigns larger weights to lower-achieving students, the mean test scores for the weighted matched sample and lower than those for the matched sample. Weighting also slightly increases the standard deviation of the 10th grade test scores but has very little impact on the correlations.

Table 3: Matched sample summary statistics

Sample	Subject	Grade 10 Aspire Plus		Grade 11 ACT		*r*
		Mean	SD	Mean	SD	
Matched Sample	Composite	202.71	24.35	20.42	5.34	0.86
	English	202.67	26.25	19.41	6.41	0.80
	Mathematics	201.26	28.62	19.94	5.22	0.77
	Reading	202.28	27.09	20.95	6.51	0.72
	Science	202.69	28.19	20.61	5.23	0.72
Matched Sample, Weighted	Composite	199.95	25.07	19.91	5.36	0.86
	English	200.01	26.83	18.87	6.41	0.80
	Mathematics	197.83	30.14	19.47	5.19	0.76
	Reading	199.92	27.63	20.50	6.51	0.72
	Science	199.29	29.05	20.13	5.25	0.72

Note: SD = standard deviation, *r* = Pearson correlation

Weighting procedure

For weighting, we used the inverse probability of treatment weight (IPTW) based on propensity scores.[2] This involves the following steps:

1) Fit a logistic regression model for the target population where the dependent variable is whether the student is included in the matched sample, and the independent variables are the demographic and achievement variables listed in Table 2.
2) Use the predicted probability from the logistic regression model as the propensity score (*ps*).
3) For students in the matched sample, assign weights as weight = 1/*ps*.

The weighted matched sample is a synthetic sample in which the distribution of covariates is independent of inclusion in the matched sample. The logistic regression model estimates used to generate the propensity scores for the composite analysis are presented in Table 4. The results are similar for the other subject areas.

The logistic regression model shows that the following variables are associated with a lower probability of being in the matched sample: Inclusion in the 1st, 2nd, or 3rd quintile of 10th grade test scores; female gender; and African American, Asian, Hispanic, or Other race/ethnicity. The weighting procedure up-weights students in these groups to make the matched sample more like the target population.

[2] Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Table 4: Logistic regression propensity score model estimates, composite score

Variable	Beta	SE	p-value
Intercept	1.075	0.027	<.001
Grade 10 Aspire Plus score, quintile*			
1st	-0.830	0.037	<.001
2nd	-0.441	0.035	<.001
3rd	-0.199	0.035	<.001
4th	-0.016	0.035	0.640
Female	-0.052	0.021	0.014
Limited English proficient	-0.047	0.053	0.376
Special education	0.045	0.039	0.250
Race/ethnicity*			
African American	-0.296	0.088	0.001
Asian	-0.643	0.072	<.001
Hispanic	-0.329	0.030	<.001
Two or more races	0.239	0.073	0.001
Other	-0.402	0.063	<.001

*Reference groups are 5th quintile and White.

Updated predicted ACT score ranges

Table 5 shows statistics related to the accuracy of the ACT score predictions and compares the accuracy of the updated predictions to those that were estimated last year. The statistics include:

- Mean width: the average width of the predicted ACT score range
- %Within: the percentage of students in the matched sample whose ACT score was within the predicted ACT score range
- %Below: the percentage of students in the matched sample whose ACT score was below the predicted ACT score range (% over-predicted)
- %Above: the percentage of students in the matched sample whose ACT score was above the predicted ACT score range (% under-predicted)

Table 5: Prediction accuracy

Subject	Previously Derived Prediction				Updated Predictions			
	Mean width	%Within	%Below	%Above	Mean width	%Within	%Below	%Above
Composite	5.5	83.0%	8.9%	8.0%	3.2	64.2%	17.2%	18.6%
English	7.9	79.2%	10.7%	10.1%	4.7	59.5%	19.4%	21.1%
Math	6.6	79.4%	16.9%	3.7%	3.4	63.8%	17.0%	19.3%
Reading	9.8	78.6%	8.6%	12.8%	5.7	57.8%	20.4%	21.8%
Science	7.2	77.7%	9.7%	12.6%	4.4	60.0%	19.8%	20.2%

Table 5 shows that:

- The updated predicted score ranges are much tighter than the previous score ranges, as shown by the decrease in the mean width of the prediction intervals.
- The updated predicted score ranges include 58-64% of actual ACT scores. Recall that the 25th and 75th conditional percentiles were used, resulting in a typical coverage percentage of around 60%. We later examine how the predicted score ranges could be widened to increase the coverage percentages.
- The updated predictions result in mostly symmetric prediction error percentages (e.g., similar percent over- and under predicted). Because the weighted matched sample has lower mean achievement than the matched sample, the updated predictions are slightly more likely to under-predict for the matched sample. (But should be more symmetric for the weighted matched sample and target population).
- While the previous predictions included a larger percentage of actual ACT scores, the predicted ranges were wider and resulted in asymmetric prediction errors (especially for math, with 16.9% over-predicted and 3.7% under-predicted).

Figure 1 below shows the updated predicted ACT score ranges (dotted green lines) as compared to those estimated last year (solid black line) for the composite scores. The figure also shows a histogram of 10th grade composite scores to show where the differences are most consequential. Figure 1 shows that the updated composite predictions generally agree with the previously-derived predictions, but with tighter prediction intervals.

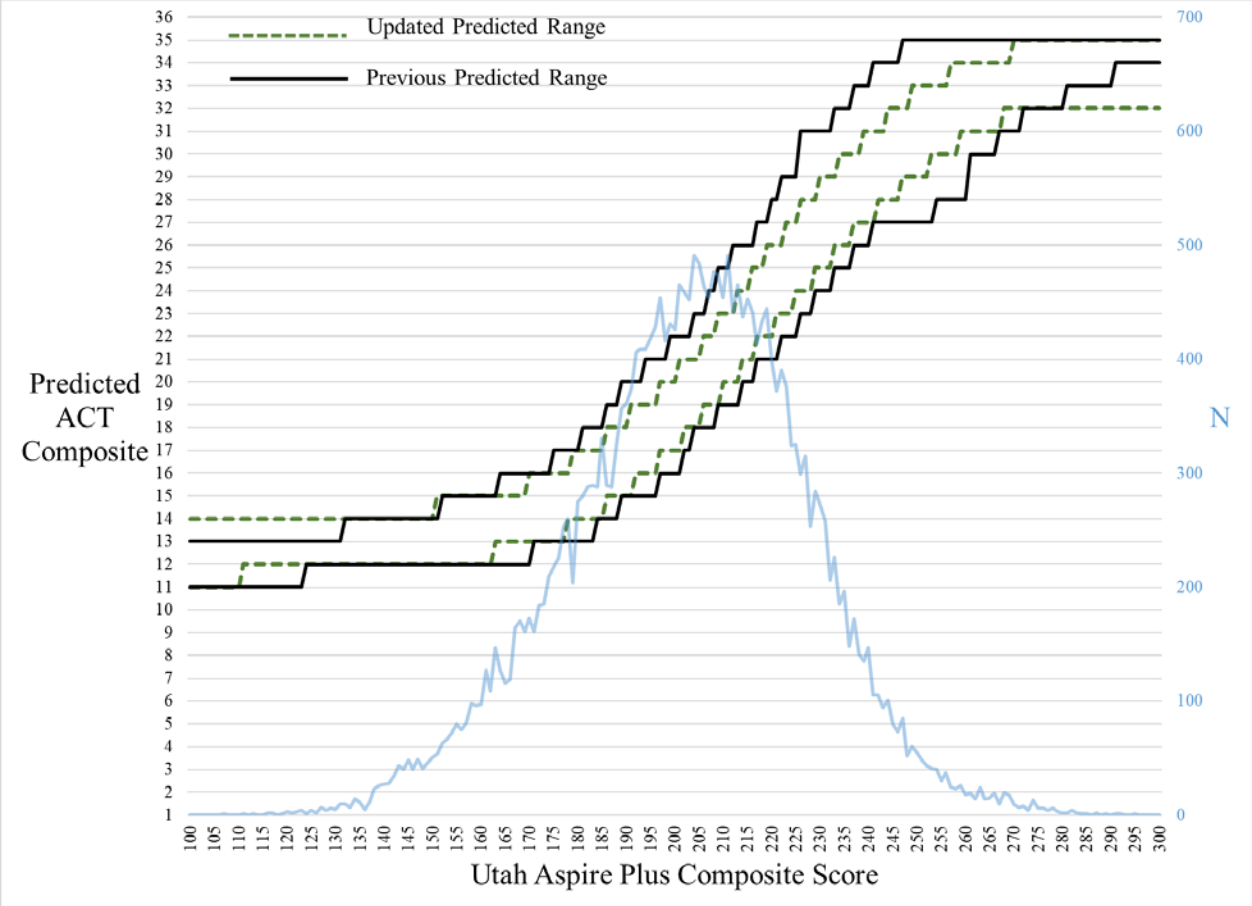

Figure 1: Predicted ACT Composite Scores

Figure 2 below shows the same information for the math scores. It shows that the math predictions changed considerably for much of the score distribution. The updated ACT math predictions have tighter intervals and are lower than those estimated last year, thus should result in fewer cases of over-prediction. Figure 2 also shows that many students (N=541) in the matched sample had the lowest possible 10th grade math score (100). It's possible that these students had a large influence on the SGP model results and updated predicted ACT score ranges. The mean ACT math score for these students was 14.6, with standard deviation 2.1. The updated predicted score range for students with a 10th grade math score of 100 (13-15) seems reasonable.

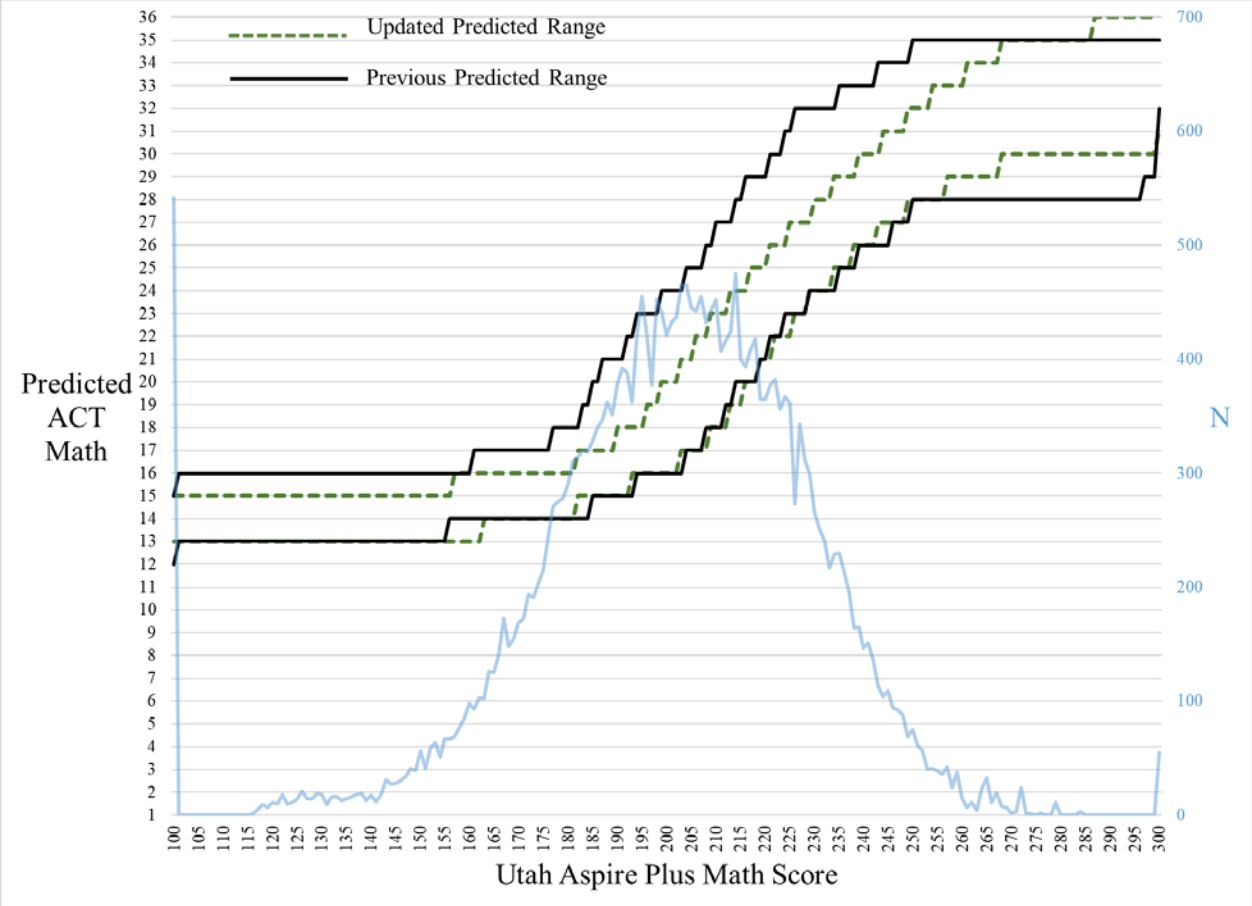

Figure 2: Predicted ACT Math Scores

Figures 1 and 2 show the results for composite and math. Comparisons of the updated and previous predicted ACT score ranges are available for the other subjects in the enclosed spreadsheet.

The predicted ACT score ranges could be widened to increase the percentage of students scoring within their predicted range. In Table 6, we explore how different conditional percentiles could be selected (instead of the 25th and 75th percentiles), resulting in wider intervals and higher percentages of students with composite scores within their predicted range.

Table 6: Impact of selecting different conditional percentiles on predicted ACT Composite score ranges

Conditional percentiles	Mean width	%Within
25th, 75th	3.2	64.2%
20th, 80th	3.9	72.2%
17th, 83rd	4.5	77.1%
15th, 85th	4.9	80.0%
10th, 90th	6.1	87.0%