STATE OF UTAH
CONTRACT AMENDMENT

AMENDMENT # 3 To CONTRACT # 136199

TO BE ATTACHED TO AND MADE A PART OF the above numbered contract by and between the State of Utah, Office of Education referred to as STATE and, American Institutes for Research, referred to as CONTRACTOR.

THE PARTIES AGREE TO AMEND THE CONTRACT AS FOLLOWS:

1. Contract period:

   02/01/2013 (original starting date)
   01/31/2018 (current ending date)
   n/a new ending date

2. Contract amount:

   $40,010,593.00 (current contract amount)
   $ 7,261,911.00 (amendment amount)
   $47,272,504.00 new contract amount add current amount to amendment amount

3. Other changes: (attach other sheets if necessary):
   Attachment C: AIR/USOE Understandings

4. Effective Date of Amendment: 09/01/2014

All other conditions and terms in the original contract and previous amendments remain the same.
IN WITNESS WHEREOF, the parties sign and cause the amendment to be executed.

CONTRACTOR

[Signature]
Contractor’s signature

[Signature]
Contract Officer
Type or Print Name and Title

STATE

[Signature]
Agency’s signature

Director, Division of Purchasing

[Signature]
Division of Finance

Janice Christensen
Agency Contact Person
801 538 7623 Telephone
801 538 7605 Fax Number
Janice.christensen@schools.utah.gov Email

Amy Bellstoe
USOE Controller

(DP-4 Revision 10/7/2009)
Attachment C: AIR/USOE Understandings

Amendment #3 to Contract 086451
American Institutes for Research/ Utah State Board of Education
September, 2014

The purpose of this contract amendment is to add to the scope of work specified in the original contract and amendments between the Utah State Office of Education (USOE) and American Institutes for Research (AIR) regarding, specifically, the addition or modification of:

- Translation from English to American Sign Language
- Human validation scoring for Writing prompts at all grades
- Lexiles
- Writing Prompt Development
- Additional Item Development

Execution of any of the tasks listed above is dependent on availability of funds. USOE reserves the right to execute any, all, or none of the elements of this contract amendment. Except as specified herein, all original and amended contract requirements, budgets, and terms & conditions remain unchanged and in effect.

Exhibit I summarizes the prices associated with the work described in this document.

Exhibit I: Summary of prices

<table>
<thead>
<tr>
<th></th>
<th>FY14-15</th>
<th>FY15-16</th>
<th>FY16-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASL Translations</td>
<td>104,200</td>
<td>1,148,496</td>
<td>1,148,496</td>
</tr>
<tr>
<td>Handscoring</td>
<td></td>
<td>1,148,496</td>
<td>1,148,496</td>
</tr>
<tr>
<td>Lexiles</td>
<td>266,000</td>
<td>170,000</td>
<td>170,000</td>
</tr>
<tr>
<td>Writing Prompts</td>
<td>3,953,481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Item Development</td>
<td>$301,238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$4,624,919</td>
<td>1,318,496</td>
<td>1,318,496</td>
</tr>
</tbody>
</table>

Translating Utah Assessment Items from English into American Sign Language

AIR shall provide ASL as an on-demand video accommodation for ELA Listening items in Utah’s computer-adaptive assessment system starting in 2014-15. Every ELA test opportunity will contain from 8 to 10 Listening items aligned to 2-3 stimuli. The stimuli will be approximately one minute each. AIR will translate 6 listening stimuli per grade to ASL, for a total of 54 translated stimuli. AIR will translate an additional 5 audio files for each grade level to be administered as embedded field test stimuli in 2014-2015.
Details

1. AIR shall develop *Guidelines for American Sign Language Translation of Assessment Items for Utah* to be used for Utah to develop summative and interim tests for implementation beginning in the 2014-2015 school year.

Translation requirements and expectations:
   a. AIR shall draft the *Guidelines* subject to USOE review and approval based on guidelines from Braille, Text-to-Speech (TTS), Foreign Language, and Smarter Balanced Signing Guidelines, including rules for items and stimuli that cannot be translated and maintain the fidelity of the measurement. AIR shall draft additional technical requirements for effective online delivery, including
      i. Video quality and specifications (e.g., frames per second [fps])
      ii. Window view parameters (e.g., zoom so that head to waist only are visible)
      iii. Background and other color requirements (e.g., clothing and jewelry)
      iv. Other requirements
   b. AIR, working with and under the direction of the USOE, shall gather input from experts and convene a committee to finalize the *Guidelines*.

2. AIR shall work with the USOE to determine how many translators to use. Translators shall meet USOE requirements, including the proper professional certification(s) and have adequate prior experience working with English-ASL translation.

3. AIR shall identify the pool of potential translators subject to USOE review and approval.

4. Translation of Training Test and Summative Test Stimuli and/or Items
   a. AIR shall provide the facility for recording and will be responsible for the actual recordings.
   b. Each item shall be translated by a certified translator.
      i. AIR shall capture multiple versions
      ii. If a translator believes a stimulus or item should be omitted, he or she will document the reason for omitting the stimulus or item.
      iii. AIR, working with and under the direction of the USOE, shall determine whether the number of items and stimuli identified for omission threatens the viability of the ASL accommodation in delivering tests that meet blueprint requirements
   c. Each translation shall be reviewed in ITS by a second USOE-approved and properly certified translator.
   d. USOE will review video as part of final web approval of the items. If USOE determines that a video is inaccurate or faulty, AIR shall re-record the video for the current administration.

5. Presentation in AIR’s Test Delivery System
   a. Each item in the ASL pool shall include the ability for students to start the ASL video with other controls similar to the TTS functionality.

6. Legal requirements
   a. AIR shall provide non-disclosure agreements
   b. AIR shall provide photo/video waivers
Exhibit 2: Development for stimuli translation:

<table>
<thead>
<tr>
<th>Grade</th>
<th># audio files/test opportunity</th>
<th>Total # audio files in bank</th>
<th>Total # audio files to be developed for year 2</th>
<th>Estimated Length</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1 minute</td>
<td>$8,800</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>45</td>
<td></td>
<td></td>
<td>$79,200</td>
</tr>
</tbody>
</table>

The total cost for year 1 and 2 of ASL translations is $79,200.

The total cost for Creation of Guidelines is $25,000.

The total cost for all ASL work is $104,200.

**Validation Hand Scoring of Writing Prompts at all Grades**

AIR shall:

- Adjust the writing item specifications to one argument/opinion essay and one informative/explanatory response in all ELA grades. This assumes two prompts per student per year.
- Provide human validation of machine-scored items beginning in 2015-2016 and continuing each subsequent contract year. The field-test planned for 2014-15 will include some machine-scored tests used for linking, and human validation of approximately 20% of those scores is included in a separate task.

During 2013-14, all Writing prompts were human scored. Scoring for 2014-15 is covered in another task. In future years, this amendment will support human validation of machine scoring. A subset of summative responses (approximately 20%) that have a higher risk of having been mis-scored will be reviewed by human scorers, based on heuristics that AIR uses to identify suspect scores. AIR will work with USOE to determine the best use of this human scoring. For example, there may be greater need for human validation at lower grades.
Note: If USOE moves forward with the Writing Field Test in the spring of 2015 as proposed in the option below, for the 2014-2015 school year, the only responses that will need to be validated are those from the fall windows. The estimated number of tests to be taken in those windows is 6,000.

**Exhibit 3: Summary of Deliverables**

<table>
<thead>
<tr>
<th>Contract year</th>
<th>Human/Hand scoring</th>
<th>Human Validation of Machine Scoring</th>
<th>Associated Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-2014</td>
<td>Yes. All writing prompts were hand scored.</td>
<td>No.</td>
<td>$2,546,625 (included in base proposal)</td>
</tr>
<tr>
<td>2014-2015</td>
<td>Yes. New prompts will be hand scored.</td>
<td>Yes. Existing, unchanged prompts used for linking will be machine-scored, 20 percent will be validated.</td>
<td>(included in writing prompts section)</td>
</tr>
<tr>
<td>2015-2016</td>
<td>Yes. New prompts will be hand scored.</td>
<td>Yes. As described in this section.</td>
<td>$1,148,496</td>
</tr>
</tbody>
</table>

Cost:
The cost for the validation hand-scoring will be $1,148,496 per year beginning 2015-16. The cost is made up of $975,000 for handsoring activities, while $173,496 is for software configuration and management oversight.

Timeline:
This will be operational in 2015-16.

**Addition of Lexiles**

Lexiles will be provided with SAGE ELA results for Grades 3 to 11. To accomplish this task, AIR shall work with MetaMetrics to provide common supplemental metrics for the measurement of reading ability. The items AIR will embed for Lexiles will go to Parent Review. There will be two distinct tasks under this agreement:

1. Scale linking. AIR shall provide:

---

1 The original contract scope included rangelining and scoring sufficient responses for the purpose of training the automated essay scoring engine during the Spring 2014 Operational Field Test. All other hand-scoring referenced in the proposal was removed.
• Lexile items for inclusion in the 2014-2015 summative test. These items shall be provided in structured, electronic format (XML), using AIR’s DTD. Licensor shall provide approximately 30-36 items per grade level, with each student answering 1-5 items
• Implementation notebooks related to Lexile Frameworks
• Utah-specific Lexile maps
• Conversion tables for SAGE scores-to-Lexile scores
• Assistance in the development of Utah-tailored Lexile websites (provide the content/links)
• Reporting on Linking studies placing SAGE reading scales on Lexiles

2. Operational Testing and Reporting, during each subsequent school year beginning 2015/16. AIR shall provide:

• Continued web-services access to websites content, including all updates provided to other clients
• Continued support required to provide Lexiles in USOE reporting

The budget for Lexiles assumes 405,000 students taking assessments each year.

Cost:
The cost to integrate Lexiles into the SAGE system for 2014-2015 is $266,000, and $170,000 for each subsequent contract year. Students in Grades 3-11 above 405,000 will incur a cost of $0.33 per student.

Timeline:
This Lexile task shall be accomplished in time for parent review in the fall of 2014 and be embedded in the spring 2015 summative window.

**Writing Prompt Development and Field Testing**

AIR will work with USOE to edit existing prompts and add an average of 8 new prompts per grade, contingent upon delivery for a late fall Content Review Committee meeting. AIR shall field test and score a total of 18 new or revised (shortened) prompts per grade, along with 6 existing prompts that will be used in their existing form to link the new prompts to the ELA scale.

The writing prompts will be developed according to the Utah SAGE Writing Task Guidelines. Writing prompts will be developed by AIR and reviewed by USOE Content staff. Once approved, a Content and Bias review will be conducted with Utah teachers reviewing all prompts. To save on costs, the Writing Bias review will be combined with the Bias review meeting scheduled for the rest of the ELA Item Development. AIR will be responsible for all costs associated with the Content and Bias review meetings. USOE Content staff will provide final direction for changes, which will be made by AIR Content staff prior to Parent Review and locking the bank.
The edited and newly developed writing prompts will be field tested starting on February 2, 2015 and ending on February 20, 2015. The main spring window for the rest of ELA testing will open on April 15, 2015. Starting on that date, students will receive real-time reports with the overall ELA score that includes writing.

At each grade, AIR shall include 12 pairs of prompts in the field test. Three of these pairs will be existing prompts in their current (long) form used to link the new prompts to the SAGE scale. The remainder will be new prompts or shortened versions of existing prompts. Existing prompts that are used for linking will be machine-scored, with 20 percent of the responses scored a second time for quality assurance purposes.

Approximately 3,300 students will respond to each of the 18 new (or modified) prompts (9 pairs). 2,000 of these responses (to be used to train the automated scoring engine) will receive two human scores, and all non-matching scores will be resolved by an expert scorer. The remainder will receive a single human score. This is summarized in Exhibit 4.

**Exhibit 4: Summary of scoring models for existing, modified, and shortened prompts**

<table>
<thead>
<tr>
<th></th>
<th>Total number</th>
<th>Number of responses double scored, 100% resolved, per prompt</th>
<th>Number of responses human scored once, per prompt</th>
<th>Number of responses machine scored with 20% human validation, per prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Prompts</td>
<td>8</td>
<td>2000</td>
<td>1335</td>
<td>0</td>
</tr>
<tr>
<td>Shortened existing prompts</td>
<td>10</td>
<td>2000</td>
<td>1335</td>
<td>0</td>
</tr>
<tr>
<td>Full length existing prompts for linking</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>3335</td>
</tr>
</tbody>
</table>

Exhibit 5 summarizes the rangefinding and scoring schedule. Prompts will go through rangefinding in late February and early March 2015.

**Exhibit 5: Writing Field Test Proposed Schedule**

<table>
<thead>
<tr>
<th>Task Name</th>
<th>Duration</th>
<th>Start</th>
<th>Finish</th>
<th>Notes</th>
<th>Resource Names</th>
</tr>
</thead>
</table>
### Milestones
- Writing window opens: 15 days
  - 1 day: Mon 2/2/15
  - 1 day: Fri 2/20/15
  - nightly responses to DRC: 15 days
  - 5 days: Mon 2/2/15
  - 10 days: Mon 2/23/15
  - 12 days: Mon 3/9/15
  - 12 days: Mon 3/16/15
  - 9 days: Fri 3/27/15

### Rangefinding and Scoring
- Rangefinding: 2 / days
  - 10 days: Mon 2/23/15
  - 10 days: Fri 3/6/15
  - 9 days: Fri 3/13/15
- DRC prepares materials: 5 days
  - 5 days: Mon 3/9/15
  - 5 days: Fri 3/13/15
- DRC scoring: 12 days
  - 12 days: Mon 3/30/15

### Preliminary file tasks
- DRC to provide preliminary file: 1 day
  - 1 day: Fri 3/30/15
- Jeff to test load with preliminary file: 2 days
  - 2 days: Sat 3/28/15
  - 2 days: Mon 3/30/15
- Extract provided to tech team: 1 day
  - 1 day: Fri 3/31/15
- Calibration tasks: 4 days
  - 4 days: Mon 4/3/15
  - 4 days: Mon 4/3/15
- Load item statistics to ITS and generate and load prelim config: 2 days
  - 2 days: Sat 4/4/15
  - 2 days: Mon 4/6/15

### Final File tasks
- Final file from DRC: 1 day
  - 1 day: Fri 4/3/15
- Final scores loaded: 2 days
  - 2 days: Mon 4/6/15
- Extract provided to tech team: 1 day
  - 1 day: Mon 4/6/15
- Calibration tasks: 5 days
  - 5 days: Fri 4/10/15
- Update ITS statistics: 1 day
  - 1 day: Fri 4/10/15
- Create config files: 1 day
  - 1 day: Sat 4/11/15
- Load config to production: 1 day
  - 1 day: Sun 4/12/15
- Re-scoring completed: 2 days
  - 2 days: Mon 4/13/15
- Spring Testing Windows: 14 days
  - 14 days: Wed 4/15/15
<table>
<thead>
<tr>
<th>Event</th>
<th>Duration</th>
<th>Date</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>math and science window opens</td>
<td>1 day</td>
<td>4/1/15</td>
<td>Wed 4/1/15</td>
</tr>
<tr>
<td>ELA window opens</td>
<td>1 day</td>
<td>4/1/15</td>
<td>Wed 4/15/15</td>
</tr>
</tbody>
</table>

Cost:

Exhibit 6 outlines the revised scope and associated cost.

**Exhibit 6: Writing Prompt Development and Field Test**
<table>
<thead>
<tr>
<th>Task</th>
<th>Scope Summary - Proposed Scope</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: New Writing Prompts</td>
<td>Develop 72 new writing prompts (8 per grade)</td>
<td>$345,519</td>
</tr>
<tr>
<td>Task 2: Edit existing prompts</td>
<td>Edit 90 existing writing prompts (10 per grade)</td>
<td>$39,229</td>
</tr>
<tr>
<td></td>
<td>-3 of the prompt pairs will be existing prompts.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Will autoscore all of the responses; second read on 20 percent of the autoscored responses.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-9 of the prompt pairs will be new prompts.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Will score each response to the new prompts twice with adjudication for 2,000 cases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Will score the remaining 1,333 cases one time with 10% second reads</td>
<td></td>
</tr>
<tr>
<td>Task 4: Rangefinding</td>
<td>Rangefinding Meeting consists of 9 committees (one for each grade) for ten days. Includes 6</td>
<td>$3,257,033</td>
</tr>
<tr>
<td></td>
<td>committee members per grade and 40 AIR staff days. Includes $100 daily stipend for participants,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>out of town participant travel for 33% of participants and meals and incidentals and facility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>costs.</td>
<td></td>
</tr>
<tr>
<td>Task 5: Content Review</td>
<td>Content Review meeting consists of 5 committees for 2.5 days. Includes 21 committee members</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and 5 AIR staff. Includes $100 daily stipend for participants, out of town participant travel</td>
<td></td>
</tr>
<tr>
<td></td>
<td>for 33% of participants, and meals and incidentals and facility costs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$25,000</td>
</tr>
<tr>
<td>Total: Writing Prompt</td>
<td></td>
<td>$3,953,481</td>
</tr>
<tr>
<td>Development &amp; Field Test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Timeline:
Writing prompt development will take place from the execution of this amendment until the field test in spring 2015.
**Revised Scope for Item Development in 2014-15**

The current contract scope for Item Development in 2014-15 includes costs for new item development (30 per grade subject), item sharing from other states (20 per grade subject), and committees, text-to-speech, and Braille.

Exhibit 7 outlines the revised scope and associated cost for the 2014-2015 Item Development

**Exhibit 7: Item Development**

<table>
<thead>
<tr>
<th>Task</th>
<th>Scope Summary - Proposed Scope</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Passages</td>
<td>Create a total of 54 passages, bring to passage review committee.</td>
<td>$68,224</td>
</tr>
<tr>
<td>Listening Stimuli</td>
<td>Create a total of 45 listening stimuli.</td>
<td>$105,244</td>
</tr>
<tr>
<td>Items</td>
<td>Review and complete a total of 1,350 teacher-drafted items (50 per grade per subject area).</td>
<td>$1,145,049</td>
</tr>
<tr>
<td>Item Writer Workshop</td>
<td>Item Writer Workshop meeting with 108 participants per subject area (ELA/Math/Science) and 75 AIR staff days. Includes $100 daily stipend for participants, out of town participant travel for 33% of participants, meals and incidentals and facility costs.</td>
<td>$591,033</td>
</tr>
<tr>
<td>Content and Fairness</td>
<td>Content and Fairness Committee meetings includes 108 committee member days and 24 AIR staff days. Includes $100 daily stipend for participants, out of town participant travel for 33% of participants, meals and incidentals and facility costs.</td>
<td>$119,688</td>
</tr>
<tr>
<td>Rubric Validation</td>
<td>Rubric validation meeting with 24 participants per subject area (ELA/Math/Science) and 54 AIR staff days. Includes $100 daily stipend for participants, out of town participant travel for 33% of participants, meals and incidentals and facility costs.</td>
<td>$140,384</td>
</tr>
<tr>
<td>TTS</td>
<td>Text-to-speech for a total of 1,350 items (50 items per grade per subject area).</td>
<td>$35,217</td>
</tr>
<tr>
<td>Braille</td>
<td>Braillng of 1,013 items (75% of items created).</td>
<td>$105,703</td>
</tr>
<tr>
<td><strong>Total: Item Development</strong></td>
<td></td>
<td>$2,310,541</td>
</tr>
<tr>
<td><strong>Original Cost</strong></td>
<td></td>
<td>$(2,009,303)</td>
</tr>
<tr>
<td><strong>Cost Increase</strong></td>
<td></td>
<td>$301,238</td>
</tr>
</tbody>
</table>

Timeline:
The additional development will be done in parallel with the Item Writer Workshops and the embedded field testing in Spring 2015.