CRESST Connected Learning Study

Novel Research Methodologies, and the Implications for PBS KIDS

Jeremy Roberts, September, 2018

All materials created that are funded by RTL needs to include the following funding language for this grant cycle: The contents of this document were developed under a cooperative agreement #PRU295A150003, from the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government. PBS KIDS® and the PBS KIDS® Logo are registered trademarks of PBS. Used with permission.
Overview

- Exploratory: Super Vision Usage
- Exploratory: Detailed Engagement Analyses
- Exploratory: Relating Specific Gameplay to Gains
- Exploratory: Detecting Misconceptions
- Implications!
Overview

- **Authors:** Elizabeth J. K. H. Redman, Gregory K. W. K. Chung, Katerina Schenke, Thomas Maierhofer, Charles B. Parks, Sandy M. Chang, Tianying Feng, Claudia S. Riveroll, and Joanne K. Michiuye
- **PI:** Gregory K. W. K. Chung
Focus on Measurement for 3-5yo: Length + height, capacity, weight
Adventure narrative
Video + Sandboxes + Games + Challenges
Modeling + Exploration + Instruction, Practice, Feedback, Assessment
Multiple operationalizations of framework constructs
Real-time monitoring of activity stream for parents
Progress and performance-based reporting
Assessment in service of learning: reporting through adults / family unit
Exploratory: Super Vision Usage

Figure 2. Parents’ usage of Super Vision.
Figure 3. Individual parents’ usage of Super Vision.
Fall off observed as you look at more and more engagement with kid learning

Wide behavior variation across parents

About a third have more robust activity pattern

We were able to see and analyze the details of what the parents did
Super Vision Usage - Implications

- Help get to the bottom of the discrepancy between parent self-report and behavior
- Potential to relate specific parent activity to learner gains
- Supports data driven approaches to entire learning ecosystem (kids, parent, teacher)
Flexibility in learner choice allows us to understand preferences.

- Average and share of time
  - by world, activity, type
- Depth of engagement
- Average number of errors
Exploratory: Detailed Engagement Analyses

- Magma Peak - Level 1
- Magma Peak - Level 2
- Slop Problem
- 12 Monkeys
- Costume Box
- Ordering Spheres
- Pirate's Tale
- Rulers
- Treasure Map
- Tree Top City - Level 1
- Tree Top City - Level 2
- Tree Top City - Level 3

Legend:
- Exit Early
- Finished
Exploratory: Detailed Engagement Analyses

- **Bird Measurer**
  - No Clicks
  - Some Activity / No Attempts
  - Some Attempts / No Successes
  - Some Successful Attempts
  - Finished Session

- **Cart Balancer**
  - No Clicks
  - Some Activity / No Attempts
  - Some Attempts / No Successes
  - Some Successful Attempts
  - Finished Session

- **Cauldron Filler**
  - No Clicks
  - Some Activity / No Attempts
  - Some Attempts / No Successes
  - Some Successful Attempts
  - Finished Session

- **Chest Sorter**
  - No Clicks
  - Some Activity / No Attempts
  - Some Attempts / No Successes
  - Some Successful Attempts
  - Finished Session

- **Mushroom Sorter**
  - No Clicks
  - Some Activity / No Attempts
  - Some Attempts / No Successes
  - Some Successful Attempts
  - Finished Session
Exploratory: Detailed Engagement Analyses

- Air Show
- All Star Sorting
- Bubble Bath
- Chow Time
- Crystals Rule
- Dino Dive
- Happy Camel
- Leaf Leader
- Pan Balance

Legend:
- No Clicks
- Some Clicks / No Attempts
- Some Attempts / No Successes
- Some Successes / No Rounds Completed
- Some Rounds Completed
- Finished Session
Exploratory: Detailed Engagement Analyses
**Exploratory: Detailed Engagement Analyses**

*Figure J1.* Share of time spent per activity type during the entire study. Every child is one observation in each boxplot.
Figure J2. Average time spent in-game per activity type. If a user exits a game and comes back to it later (or immediately) the visits are counted separately. Every observation is the average time a user spends in a title of a certain activity type.
Exploratory: Detailed Engagement Analyses

Figure J3. Total number of visits per activity type.
Figure J4. Total time spent per child per activity, for every activity.
Figure J6. Average time spent in-game per activity for every activity.
Figure J7. Average time spent in-game per activity for every child.
Exploratory: Detailed Engagement Analyses

Figure J8. Number of visits per activity.
Figure J9. Average time spent in an activity by world per day (1-12).
Figure J10. Share of time spent in each world per day.
Exploratory: Detailed Engagement Analyses

*Figure J11.* Share of time spent on activity types over the course of the study.
Figure J12. Average time spent in-game by activity type.
Exploratory: Detailed Engagement Analyses

Figure J13. Time spent in-game by activity type.
Exploratory: Detailed Engagement Analyses

Figure J14. Share of time spent per activity over the course of the study.
Exploratory: Detailed Engagement Analyses

Figure J15. Average time spent in-game.
Exploratory: Detailed Engagement Analyses

Figure J17. The average number of distinct titles entered per kid per day per activity.
Engagement Analyses: Implications

- We can use PBS KIDS LAP to help understand how content actually gets used (including patterns over time)
- We are now piloting this kind of analyses “on the fly” during product design and iteration
- There is a wide variation in favorite content
- We can see how far kids are getting, and how deeply they are engaging.
- We can identify under and out performers
Engagement Analyses: Implications

- We can understand initial appeal, return appeal, frequency, staying power
- We can start to get at the dosage and timescales involved in learning
- We can compare and compete design strategies
- We can understand contributions made by individual components in media aggregation strategies
- Stickers rule!
## Exploratory: Relating Gameplay to Gains

**Table 10**

*Coefficients for Hierarchical Linear Regression Models Predicting Posttest Score Based on Pretest Score, Treatment Condition, and Site*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
<th>Q12</th>
<th>Q13</th>
<th>Q14</th>
<th>Q15</th>
<th>Q16</th>
<th>Q17</th>
<th>Q18</th>
<th>Q19</th>
<th>Q20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.41</td>
<td>.44</td>
<td>.51</td>
<td>.32</td>
<td>.35</td>
<td>.45</td>
<td>.24</td>
<td>.31</td>
<td>.50</td>
<td>.39</td>
<td>.23</td>
<td>.24</td>
<td>.29</td>
<td>.32</td>
<td>.08</td>
<td>.21</td>
<td>.19</td>
<td>.49</td>
<td>.25</td>
<td>.07</td>
</tr>
<tr>
<td>Pretest</td>
<td>.54</td>
<td>.31</td>
<td>.36</td>
<td>.60</td>
<td>.13</td>
<td>.40</td>
<td>.61</td>
<td>.29</td>
<td>.42</td>
<td>.33</td>
<td>.46</td>
<td>.51</td>
<td>.47</td>
<td>.40</td>
<td>.43</td>
<td>.33</td>
<td>.37</td>
<td>.48</td>
<td>.59</td>
<td>.81</td>
</tr>
<tr>
<td>MU+SV</td>
<td>.07</td>
<td>.09</td>
<td>.07</td>
<td>.05</td>
<td>.14</td>
<td>.04</td>
<td>-.03</td>
<td>.22</td>
<td>-.06</td>
<td>.03</td>
<td>.16</td>
<td>-.01</td>
<td>.13</td>
<td>.20</td>
<td>.14</td>
<td>.27</td>
<td>.35</td>
<td>.04</td>
<td>.05</td>
<td>-.05</td>
</tr>
<tr>
<td>MU</td>
<td>.04</td>
<td>.04</td>
<td>.02</td>
<td>.01</td>
<td>.27</td>
<td>.04</td>
<td>-.02</td>
<td>.11</td>
<td>.02</td>
<td>.25</td>
<td>.02</td>
<td>.01</td>
<td>.09</td>
<td>.19</td>
<td>.22</td>
<td>.19</td>
<td>.20</td>
<td>-.02</td>
<td>-.05</td>
<td>.06</td>
</tr>
<tr>
<td>Site 1</td>
<td>-.02</td>
<td>.02</td>
<td>.01</td>
<td>.00</td>
<td>-.05</td>
<td>-.01</td>
<td>-.03</td>
<td>-.06</td>
<td>-.03</td>
<td>.09</td>
<td>-.01</td>
<td>.01</td>
<td>.00</td>
<td>-.02</td>
<td>-.03</td>
<td>-.03</td>
<td>.00</td>
<td>-.02</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Site 2</td>
<td>-.03</td>
<td>-.08</td>
<td>-.01</td>
<td>.00</td>
<td>-.05</td>
<td>-.04</td>
<td>-.10</td>
<td>-.12</td>
<td>-.05</td>
<td>.02</td>
<td>-.04</td>
<td>-.03</td>
<td>.00</td>
<td>-.04</td>
<td>-.06</td>
<td>-.02</td>
<td>.00</td>
<td>-.06</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Site 3</td>
<td>-.04</td>
<td>.05</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.02</td>
<td>.10</td>
<td>.15</td>
<td>.02</td>
<td>-.07</td>
<td>-.03</td>
<td>.03</td>
<td>.00</td>
<td>.02</td>
<td>.08</td>
<td>.02</td>
<td>.00</td>
<td>.08</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Site 4</td>
<td>.09</td>
<td>.06</td>
<td>.01</td>
<td>.00</td>
<td>.18</td>
<td>.05</td>
<td>.14</td>
<td>.10</td>
<td>.13</td>
<td>-.01</td>
<td>.10</td>
<td>.03</td>
<td>.00</td>
<td>.02</td>
<td>.08</td>
<td>.10</td>
<td>.00</td>
<td>.11</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Site 5</td>
<td>.01</td>
<td>-.05</td>
<td>-.02</td>
<td>.00</td>
<td>-.08</td>
<td>-.01</td>
<td>-.11</td>
<td>-.08</td>
<td>-.07</td>
<td>-.02</td>
<td>-.02</td>
<td>-.03</td>
<td>.00</td>
<td>.02</td>
<td>-.08</td>
<td>-.06</td>
<td>.00</td>
<td>-.11</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>
# Exploratory: Relating Gameplay to Gains

Table 11

*Coefficients for Hierarchical Linear Regression Models Predicting Question Categories on Posttest Score Based on Pretest Score, Treatment Condition, and Site*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Capacity (Q2-3)</th>
<th>Length (Q4-5, 8-10, 20)</th>
<th>Displacement (Q11)</th>
<th>Height (Q6-7, 18-19)</th>
<th>Weight (Q12-13)</th>
<th>WeightIPad (Q14-17)</th>
<th>NolIPad (Q1-13, 18-20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.41</td>
<td>.21</td>
<td>.23</td>
<td>.38</td>
<td>.22</td>
<td>.16</td>
<td>.54</td>
</tr>
<tr>
<td>Pretest</td>
<td>.43</td>
<td>.61</td>
<td>.46</td>
<td>.58</td>
<td>.60</td>
<td>.53</td>
<td>.20</td>
</tr>
<tr>
<td>MU+SV</td>
<td>.08</td>
<td>.06</td>
<td>.16</td>
<td>.03</td>
<td>.05</td>
<td>.22</td>
<td>.05</td>
</tr>
<tr>
<td>MU</td>
<td>.03</td>
<td>.13</td>
<td>.02</td>
<td>-.01</td>
<td>.05</td>
<td>.21</td>
<td>.06</td>
</tr>
<tr>
<td>Site 1</td>
<td>.00</td>
<td>-.01</td>
<td>.09</td>
<td>-.02</td>
<td>.00</td>
<td>-.01</td>
<td>-.05</td>
</tr>
<tr>
<td>Site 2</td>
<td>-.07</td>
<td>-.04</td>
<td>.02</td>
<td>-.02</td>
<td>-.04</td>
<td>-.04</td>
<td>-.07</td>
</tr>
<tr>
<td>Site 3</td>
<td>-.01</td>
<td>.04</td>
<td>-.07</td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Site 4</td>
<td>.09</td>
<td>.06</td>
<td>-.01</td>
<td>.08</td>
<td>.07</td>
<td>.03</td>
<td>.14</td>
</tr>
<tr>
<td>Site 5</td>
<td>-.01</td>
<td>-.05</td>
<td>-.02</td>
<td>-.05</td>
<td>-.05</td>
<td>-.01</td>
<td>-.06</td>
</tr>
</tbody>
</table>

*Note. The questions falling into the specific category are in parentheses.*
# Exploratory: Relating Gameplay to Gains

## Table 13

*Child Test and Gameplay Data Related to Weight Domain*

<table>
<thead>
<tr>
<th>Child ID</th>
<th>Pretest scores on weight items</th>
<th>Posttest scores on weight items</th>
<th>Score difference on pre- and posttest</th>
<th>Question 12 pretest and posttest responses</th>
<th>Most time spent on Crystal Caves activities</th>
</tr>
</thead>
</table>
| 116      | 3.0                           | 5.0                             | 2                                    | *Pre:* Goes up down and side to side and spins. Swings.  
*Post:* They're pan balances. You use something on it to see which one is heavier. The side that tips down is heavier. | • Pan Balance (20 min)  
• Chow Time (19 min)  
• Happy Camel (17 min) |
| 121      | 2.0                           | 5.0                             | 3                                    | *Pre:* (no response)  
*Post:* This one measures stuff. How heavy is something. | • Leaf Leader (19 min)  
• Pan Balance (18 min) |
| 206      | 2.0                           | 6.0                             | 4                                    | *Pre:* I don't know.  
*Post:* Pan balance. Balance things. How heavy is it. | • Chow Time (24 min)  
• Leaf Leader (20 min) |
| 212      | 0.0                           | 5.0                             | 5                                    | *Pre:* They do of the same things and then they go down.  
*Post:* They put things so they can weigh and tell if they're balanced | • Balancing Act (22 min) |
| 213      | 2.5                           | 4.5                             | 2                                    | *Pre:* They measure.  
*Post:* Measure. | • Pan Balance (30 min) |
| 225      | 0.5                           | 4.5                             | 4                                    | *Pre:* They hold how many weight it takes.  
*Post:* They hold up stuff that's the most weight. | • Cart Balancing (11 min) |
Relating Gameplay to Gains: Implications

- We can know what items showed the highest gains
- We can find the children who showed gains on those items
- We can examine those kids' telemetry
- We can connect the dots!
• Metz 1993: Preschoolers and Pan Balances
## Exploratory: Detecting Misconceptions

### Table K1

**Strategies: Substance and Organization (From Metz, 1993, Table 2)**

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-to-one correspondence</td>
<td>The subject puts one element in each of the pans, by pairs or alternation of placements. Action is complete when all elements have been distributed. Focus on completion of the action (exhaustive distribution) and not developing cardinal number.</td>
</tr>
<tr>
<td>Equal number</td>
<td>The subject puts the same number in each pan. Concern with evolving quantity in each pan, as manifested by preliminary halving of the set or counting of set contents (as indicated by finger or verbal tagging).</td>
</tr>
<tr>
<td>Visual feedback</td>
<td>The subject determines next placement by apparatus's response resulting from previous placement, as indicated by train of visual attention, rhythm and sequence of actions on the apparatus, and verbal protocol.</td>
</tr>
<tr>
<td>Equal weight</td>
<td>The subject directly (not step-by-step by visual feedback) divides the set into two subsets of equal weight, using equal number if and only if confident the elements weigh the same and, when weights vary, by compensations between number of elements and their respective weights. Indicated by prearrangements of the elements on the table, train of visual attention, rhythm and sequence of placements in the pans, and verbal protocol.</td>
</tr>
<tr>
<td>Knob fiddling</td>
<td>The subject adjusts the knobs built into the beam. (Given the light weight of the knobs, knob fiddling never constitutes successful repair. In addition, directions of fiddling were also coded to differentiate from &quot;other: torque adjustment.&quot;)</td>
</tr>
<tr>
<td>Compensation by force of hands</td>
<td>The subject tries to attain the goal state by forcing the apparatus into the goal state or by yanking down or thrusting up on the beam, the wire, or the pan. Indicated by actions on the apparatus and corresponding expectations regarding its conservation (as inferred from facial expression, subsequent actions, and verbal protocol).</td>
</tr>
<tr>
<td>Realign</td>
<td>The subject tries to attain the goal state by grasping the beam, the wire, or the pan, bringing into goal alignment, and gently releasing. Indicated by actions on the apparatus and corresponding expectations regarding its conservation (as inferred from facial expression, subsequent actions, and verbal protocol).</td>
</tr>
<tr>
<td>Exchange</td>
<td>The subject switches pan placements of one element from each pan or switches whole pan contents from one side to the other. Indicated by rhythm and sequence of manipulations in relation to examination of the apparatus's response and verbal protocol.</td>
</tr>
<tr>
<td>Reject</td>
<td>The subject excludes an element from placement in either pan. Indicated by handing the experimenter one or more elements, placing an element(s) aside, or verbal request (disallowed).</td>
</tr>
</tbody>
</table>

### Table K2

**Representation: Apparatus Focus (From Metz, 1993, Table 4)**

<table>
<thead>
<tr>
<th>State</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Yank or align</td>
<td>The subject yanks or aligns the apparatus for the purpose of attaining the goal, as evidenced by rhythm and sequence of the subject's apparatus manipulations, verbal protocol, and facial expressions as reflection of expectation or surprise.</td>
</tr>
<tr>
<td>2. Displace elements across pans</td>
<td>The subject attempts to attain the goal state either by moving elements from the up-pan to the down-pan or by moving elements in both directions, as evidenced by attention to the tilt of the apparatus after element placements, rhythm and sequence of examining tilt of apparatus and element displacements, directions of displacements, and verbal protocol. Absence of concern with heavier element per se, as evidenced by train of visual attention, manipulations, and verbal protocol.</td>
</tr>
<tr>
<td>3. Move elements from down-pan to up-pan: visual feedback</td>
<td>The subject attempts to attain the goal state by displacing elements from the down-pan to the up-pan, as evidenced by attention to the tilt of the apparatus after element placements, rhythms and sequence of examining tilt of apparatus and element displacements, directions of displacements, and verbal protocol. Absence of concern with heavier element per se, as evidenced by train of visual attention, manipulations, and verbal protocol.</td>
</tr>
</tbody>
</table>
## Exploratory: Detecting Misconceptions

### Table K3

**Representation: Weight Focus (From Metz, 1993, Table 4)**

<table>
<thead>
<tr>
<th>State</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No representation of the elements as weights</td>
<td>The subject represents neither the weight differences between the elements nor the element as a weight, as evidenced by absence of differential attention in examination of the elements, absence of differential treatment in action, and verbal protocol.</td>
</tr>
<tr>
<td>2. Representation of weight without relevance</td>
<td>The subject represents the elements as weights but does not link the property to the task at hand. Representation is evidenced by differential examination of the elements, with the more intense examination given to the heavy element, and by verbal protocol. Lack of relevance is evidenced by verbal protocol and by absence of differential treatment in action.</td>
</tr>
<tr>
<td>3. Relevance of weight without specific implications</td>
<td>The subject believes that the weight differential is somehow related to the task but has not yet elaborated any specific implications. Relevance is evidenced by differential attention in examination of the elements, with the more intense examination given to the heavy element, and by verbal protocol. Absence of implications is evidenced by failure to base either diagnosis of the disequilibrium or action on the differential weights.</td>
</tr>
<tr>
<td>4. Diagnostic implication</td>
<td>The subject interprets the tilt of the apparatus, under conditions of equal number distributions, as signifying the placement of a heavy element in the lower pan, but weights-based thinking does not yet extend to weights-based action. Evidenced by interactions with the apparatus and verbal protocol.</td>
</tr>
<tr>
<td>5. Search for an instrumental implication</td>
<td>The subject searches for an instrumental implication to resolve the “one element weighs too much” problem, as evidenced by weight-related but inappropriate actions such as exchanging placements between a heavy weight and a standard weight, trying to reject a heavy weight (disallowed), or reshaping a heavy weight. Evidenced by nature of the actions carried out on different weights and verbal protocol.</td>
</tr>
<tr>
<td>6. Repair implication: visual feedback</td>
<td>The subject displaces weights from the lower pan to the raised pan or places the next element from the pile into the raised pan as a solution to the “one element weighs too much” problem, as evidenced by rhythm and sequence of interactions with the apparatus and different elements and by verbal protocol.</td>
</tr>
<tr>
<td>7. State implication: compensation</td>
<td>The subject solves the “one element weighs too much” problem by trade-offs between number of elements in the two pans and their respective weights, as evidenced by the decision of pan collections prior to any placements or feedback from the apparatus, the formation of states with number versus weights trade-offs, and verbal protocol.</td>
</tr>
</tbody>
</table>
Exploratory: Detecting Misconceptions

Figure K1. Interaction space mapping for “Pan Balance.”
Detecting Misconceptions
Detecting Misconceptions

**STUDENT BKGD LAYER**
- Prior knowledge, game experience
- Age, sex, language proficiency

**CONSTRUCT LAYER**
Construct, subordinate constructs, and inter-dependencies

**INDICATOR LAYER**
Behavioral evidence of construct,
\[ o_n = f_n(q_1, q_2, q_3, \ldots, q_n) \]

**TRANSFORMATION FUNCTIONS**
Extracts indicators from raw telemetry

**EVENT LAYER**
Raw player behavior and game states

*Figure K5. Computational structure.*
## Detecting Misconceptions

**Table K4**

*Descriptive Statistics of Misconception and Weight Measures*

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher is heavier misconception</td>
<td>53</td>
<td>1.64</td>
<td>3.13</td>
<td>0.0</td>
<td>18.00</td>
</tr>
<tr>
<td>Yank or align misconception</td>
<td>53</td>
<td>3.49</td>
<td>4.87</td>
<td>0.0</td>
<td>21.00</td>
</tr>
<tr>
<td>Pretest – weight items</td>
<td>66</td>
<td>2.05</td>
<td>1.58</td>
<td>0.0</td>
<td>6.00</td>
</tr>
<tr>
<td>Posttest – weight items</td>
<td>66</td>
<td>3.18</td>
<td>1.87</td>
<td>0.0</td>
<td>6.00</td>
</tr>
<tr>
<td>Gain (Posttest – Pretest) – weight items</td>
<td>66</td>
<td>1.13</td>
<td>1.59</td>
<td>-2.0</td>
<td>5.00</td>
</tr>
<tr>
<td>Time spent on “Pan Balance” (sec)</td>
<td>66</td>
<td>312.95</td>
<td>371.01</td>
<td>0.0</td>
<td>1792.17</td>
</tr>
<tr>
<td>Time spent on “Cart Balancer” (sec)</td>
<td>66</td>
<td>169.46</td>
<td>140.90</td>
<td>0.0</td>
<td>680.88</td>
</tr>
</tbody>
</table>
### Detecting Misconceptions

**Table K5**

*Intercorrelations Among Misconception Measures and External Measures of Weight*

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Higher is heavier misconception(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Yank or align misconception(^a)</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Pretest – weight items(^b)</td>
<td>-.04</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Posttest – weight items(^b)</td>
<td>-.36**</td>
<td>-.23</td>
<td>.58***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Gain (Posttest – Pretest) – weight items(^b)</td>
<td>-.34*</td>
<td>-.19</td>
<td>-.30*</td>
<td>.59***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Time spent on “Pan Balance”(^b)</td>
<td>.21</td>
<td>.08</td>
<td>.15</td>
<td>.18</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>7. Time spent on “Cart Balancer”(^b)</td>
<td>.13</td>
<td>-.14</td>
<td>.15</td>
<td>.11</td>
<td>-.02</td>
<td>.33**</td>
</tr>
</tbody>
</table>

\(^a\)\(n = 53\). \(^b\)\(n = 66\).

*\(p < .05\), two-tailed. **\(p < .01\), two-tailed. ***\(p < .001\), two-tailed.*
Detecting Misconceptions: Implications

● We can consult the literature for established misconceptions
● With the right interactives, we can detect them
● We can relate having a misconception to gains, or lack thereof
● Great foundation for PAL: provide support tailored to a misconception
Overall Implications for the Future

- We now have a great methodology for evaluation
- We can use PBS KIDS LAP for unprecedented understanding of how products and services get used
- We can connect the dots between learning gains, and the specific engagement activity that correlates to it
- We can consult literature for well-defined misconceptions, and detect them
- We think we can use knowledge of misconceptions to help power Personalized and Adaptive Learning efforts
Thank You!
Thank You!
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

- Director: Li Cai  |  Founding director: Eva Baker

Expertise in the application of technology to measure knowledge and skills in low-stakes settings
- Knowledge representation (ontologies)
- Simulations and games (design, automated scoring)
- Process data, evaluation of learning technologies

Focus on validity
- How does cognition and motivation manifest itself in digital environments, and what are behavioral indicators of both?
- Use of process data to support measurement of complex performance
Technology R&D

- Intersection of learning, instruction, measurement, and assessment
- Education, military, ed tech
  - PBS KIDS, EDC, NMSI, ETS
  - IES, NSF, Office of Naval Research, DARPA, USN
- PreK to adults
- Effectiveness studies (RCT / adhere to WWC standards)
- Exploratory studies
Goal: Examine the effects of MU and SV on children’s learning of measurement concepts and parents’ awareness and support of their children’s mathematics learning
Pretest-posttest randomized design

Three conditions:
- PBS KIDS Measure Up! (MU)
- MU + PBS KIDS Super Vision (SV)
- Control (Literacy apps: Super WHY!)

Connected Learning Study
Sample

- 4-5 years old
- Public elementary schools (Title I)
  - Prekindergarten and transitional kindergarten classrooms
- Childcare center
  - Affiliated with community college
Protocol

● Four days a week, three weeks
● 20 - 30 minutes a day
● Max 6 hours
● In-classroom play
Assessments

- 20 items, 10 minutes, 1-on-1
  - Length and height
  - Capacity and displacement
  - Weight
Assessments

- Adopted, adapted, and created external criterion measures
- Four CMA items
  - Extended CMA with author’s support
- Three KeyMath-3 items
- CRESST created 13 items
Assessments

- Eleven employed manipulatives
- Six used pictures
- Four utilized an iPad app
  - Created by CRESST to simulate using a pan balance
- Usability testing and expert review
Assessments
Assessments

A.

B.

C.

D.
And More

- Lots more available to view in report
  - All protocols, assessment items, and surveys
  - Background and demographic info
  - Descriptive statistics
  - Model definitions
Four models
- Each introduces a different covariate
  - Site, age, gender, low-income
  - Robust and consistent across all 4 models

Predictors explain a substantial amount of variance in the outcome variable
MU condition results in **more than two** additional correct responses on posttest
MU + SV results in **more than 1.5** additional correct responses on posttest
Statistically significant gains compared to control (all between 8-11% improvement)
Not statistically different from one another