Formative Assessment Delivery System

William Fisher, Research Associate
Perman Gochyyev, Graduate Researcher
Mark Wilson, Principal Investigator

FADS

The project aimed to develop a computerized system to:

- Design
- Develop
- Deliver
- Report

on assessments that help teachers diagnose students' comprehension and learning needs.

Why FADS?

- **learning progressions**
- **complex item formats** that can take advantage of rich multimedia platforms
- **sophisticated modeling techniques** for robust evidence

BEAR Assessment System Principles

Developmental Perspective	Need a framework for communicating meaning
Match between Instruction and Assessment	Need methods of gathering data that are acceptable and useful to all participants
Interpretable by Teachers	Need a way to value what we see in student work
Evidence of Quality	Need a technique of interpreting data that allows meaningful reporting to multiple audiences

"Full-cycle" production

- Measurement framework:
  - Four building blocks:
    - the *construct map*,
    - the *items design*,
    - the *outcome space*,
    - the *measurement model*.
  - Activities from hypothesizing about the construct to be measured to making interpretations and decisions – can be organised into these four building blocks.

CDMW

- The Constructing Data, Modeling Worlds (CDMW)
  - aims to investigate the development of students' skills and knowledge related to data modeling and statistical reasoning in elementary and middle schools.
- In coordination with teachers using the *Constructing Data, Modeling Worlds* (CDMW*)* curriculum, developed at Vanderbilt University.

CDMW: Constructs

- Chance
- Conceptions of Statistics
- Data Display
- Informal Inference
- Modeling Variability
- Meta Representational Competence
- Theory of Measurement

Construct: Data Display

	Level		Performances	Examples
DaD6	Integrate case with aggregate perspectives.	*DaD6* A	Discuss how general patterns or trends are either exemplified or missing from subsets of cases.	▪ Relate qualities of a case as an example of general qualities of a region of data (case as typical of data region). ▪ Notice that a subset of cases does not seem to fit the trends observed or conjectured.
DaD5	Consider the data in aggregate when interpreting or creating displays.	*DaD5* B	Quantify aggregate property of the display using one or more of the following: ratio, proportion or percent.	▪ "I found out that measurements between 45 and 55 were 70% of our measurements. So, I guess the true height is somewhere between 45 and 55." ▪ Students annotate their display to show percentages within particular regions.
		DaD5 A	Recognize that a display provides information about the data as a collective.	▪ "The distribution of the data is wider for rounded-nosecone rockets than for pointed-nosecone rockets. Maybe that's because pointed rockets flights are more consistent." ▪ "When we measure different things, we keep getting a bell shape. That's because we tend to get around the real measure most of the time, but sometimes we make big mistakes."

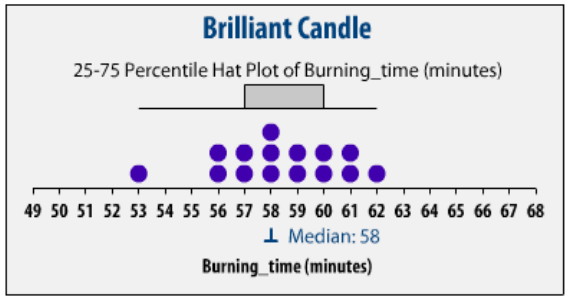
Construct: Data Display

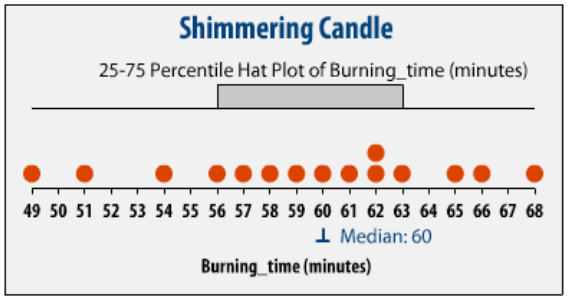
DaD3	**Notice or construct groups of similar values.**	*DaD3* **A**	Notice or construct groups of similar values from distinct values.	▪ Create unordered bins, and comment on, for example, the number of occurrences of 40s vs. the number of 50s. ▪ When asked to name bins in a preset display, assigns discontinuous and/or unequal intervals to the bins, such as 2-25, 26-36, 37-45. ▪ Create equal interval bins but leave out intermediate intervals. ▪ Notice "plateaus" in the case display or a group of similar values. ▪ "This number, 193, is really different, because the others are all between 160 and 165." ▪ "Most of batteries lasted between 120 to 140 minutes."
DaD2	**Interpret and/or produce data displays as all collections of individual cases.**	*DaD2* **B**	Construct/interpret data by considering ordinal properties.	▪ "The data start out with the lowest measurement and go to the highest one." ▪ Create display by ordering data as a list or case-value graph.
		DaD2 **A**	Concentrate on specific data points without relating these to any structure in the data.	▪ Identify maximum and minimum values. ▪ "The only thing I can tell is this (193) is the highest." ▪ "154 is the number in the middle of the list (without ordering the data)." ▪ "This number is the biggest."
DaD1	**Create displays or interpret displays without reference to goals of data creation.**	*DaD1* **A**	Create or interpret data displays without relating to the goals of the inquiry.	▪ "We grouped even and odd numbers because we like even and odd numbers." ▪ "I put these two values (19 and 11) on the top because that's my birthday - Nov. 19th!" ▪ "This display has lots of numbers."

examples of the developed items

Brilliant Candle vs. Shimmering Candle

Shimmering Candle Company claims that their candles burn longer, on average, than candles made by the Brilliant Candle Company. Testers for Consumer News burned 15 Shimmering candles and 15 Brilliant candles and recorded the number of minutes that each candle burned. The plot below shows the burning time and the middle-50 percent of the data around the median.



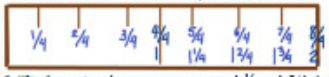
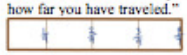


- Is the claim made by the Shimmering Candle Company, that their candles burn longer, supported by the test results?

  ○ Yes ○ No

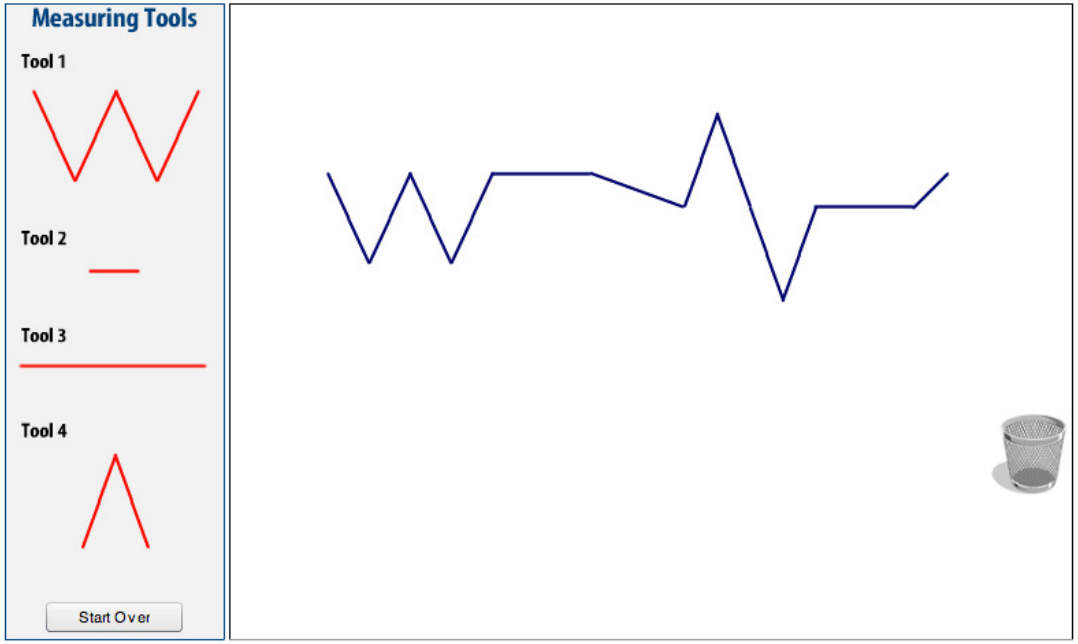
- Why do you think so?

Construct: Theory of Measurement

	Level	Performances		Examples
ToM 7	Use theory of measurement to solve novel problems.	*ToM7* B	Invent and justify measure.	• "This bushiness index (extent of branching) tells me how much the elodea plant (its total length) grew in the water."
		ToM7 A	Use derived units to solve measurement problems.	• "We can use the time it takes as a measure of distance, because we can assume a constant rate."
ToM 6	Predict the effects of changes in unit on measure or on scale.	*ToM6* C	Evaluate tradeoffs when selecting methods and tools for measurement.	• We could measure the width of the classroom with our 15 cm. ruler, but it would take forever. And we don't need to be so accurate to decide how to re-arrange our desks *(not an actual student response).*
		ToM6 B	Derive relations among units, given expression of the same attribute in different scales of measure.	• "If the measure of the height of the plant is about 10 cm or about 4 inches, then an inch is about 2 ½ cm."
		ToM6 A	Use relations among units to quantify results of changes in unit.	• "The measure of the height of the plant is 14 cm. If a cm. is 10 times as long as a mm, then the measure is 140 mm." • "If I change the unit so that it is half as long as the original unit, the measure doubles."
ToM 5	Coordinate units to constitute a scale.	*ToM5* D	Account for change of origin when measurement does not start at zero. *(Zero-Point: Any number can serve as zero.)*	• "I can start to measure from the 3 on the inches ruler, and take off 3 inches from the result." (Note: Starting at 2 or 3 is generally more difficult than starting at 1. Starting at a non-whole number is more difficult still). • "If I start at 3 and go to 7 ¼, the measure is 4 1/4." (Note: Shifts involving fractions are usually more difficult than those with whole numbers.)
		ToM5 C	Compose units and represent the composition on a scale.	• "It was 15 inches or 1 ¼ feet." *Note:* These are 2-level units, or units-of-units.  • "Its 6 quarters long, or we can say 1 ½ or 1 2/4 Anyia long (the name of the nonstandard unit).
		ToM5 B	Generate and compose splits of units involving odd factors, such as 3 or 5.	• "I made 12th's like this. First I split the unit into 3; then I split it by 2 and then 2 again. That makes 12th's." • "A meter is 100 times as long as a cm." • "A cm is 1/100 times as long as a m."
		ToM5 A	Symbolize relation between origin and partitioned units on scale.	• "You don't write 1/4 in the middle. . .  • . . .Put it at the end of the part of the unit, so you can see how far you have traveled." 

Ant Path, Continued

- Select one tool, then drag onto the stage as many copies of it as you need to measure the length of the path.



- Using the tool you chose as a unit of measurement, approximately how long is the path?

The path measures ☐

FADS Research Question 1

- *To what extent can responses to formative assessment activities developed from selected existing mathematics and science curricula be mapped to progress variables and standards?*

FADS Research Question 2

- *To what extent can these formative assessment tasks in science and mathematics be delivered and scored by computer? Where automated scoring is not possible, how can teachers be assisted in scoring responses?*

FADS Research Question 3

- *How well do the graphical and narrative reports reflect the instructional needs of students? How useful do teachers find the reports? In what ways do teacher appreciation and use of formative assessment data change when such data become more readily available and/or easier to understand?*

FADS Research Question 4

- *Do students find computerized delivery of assessment activities more engaging than paper-and-pencil versions? Do they initiate these activities more often when they are available via technology? Are there other benefits from computerized interaction or feedback found in the computerized versions?*

CDMW beta trials

- The study can be divided into four major milestones for the participating teachers
  - (a) a training session,
  - (b) the first online assessment,
  - (c) the second online assessment,
  - (d) a closing interview.

CDMW beta trials

(a) a training session

(b) the first online assessment

(c) the second online assessment

(d) a closing interview

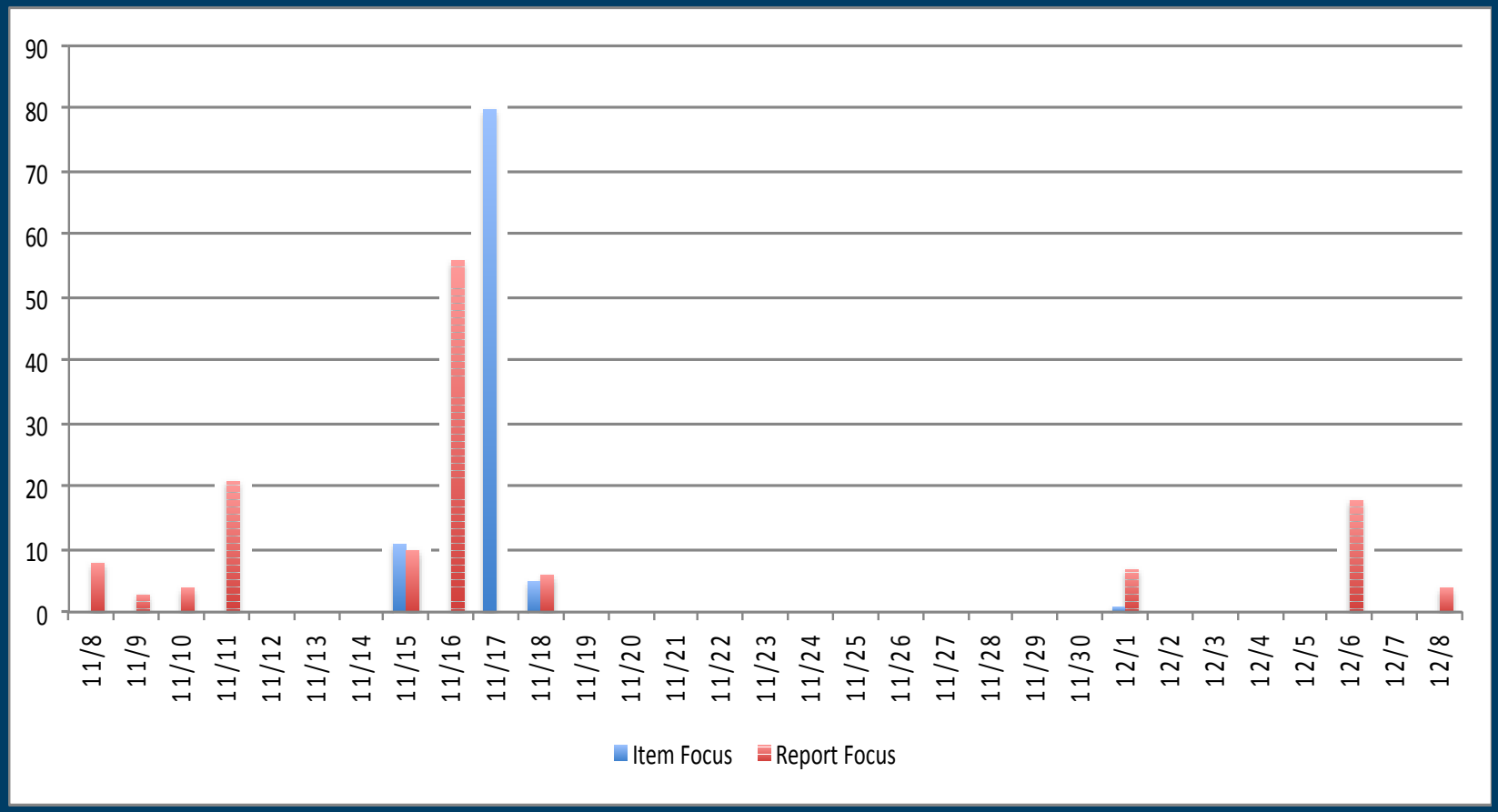
Training session

- Introduced the participating teachers to the use of FADS and the resources that it offers.
- Teachers were randomly assigned to a given training session.
- **Goal**: to inform how alternative emphases during training influenced access to and use of the formative assessment report resources offered by FADS.

Training session

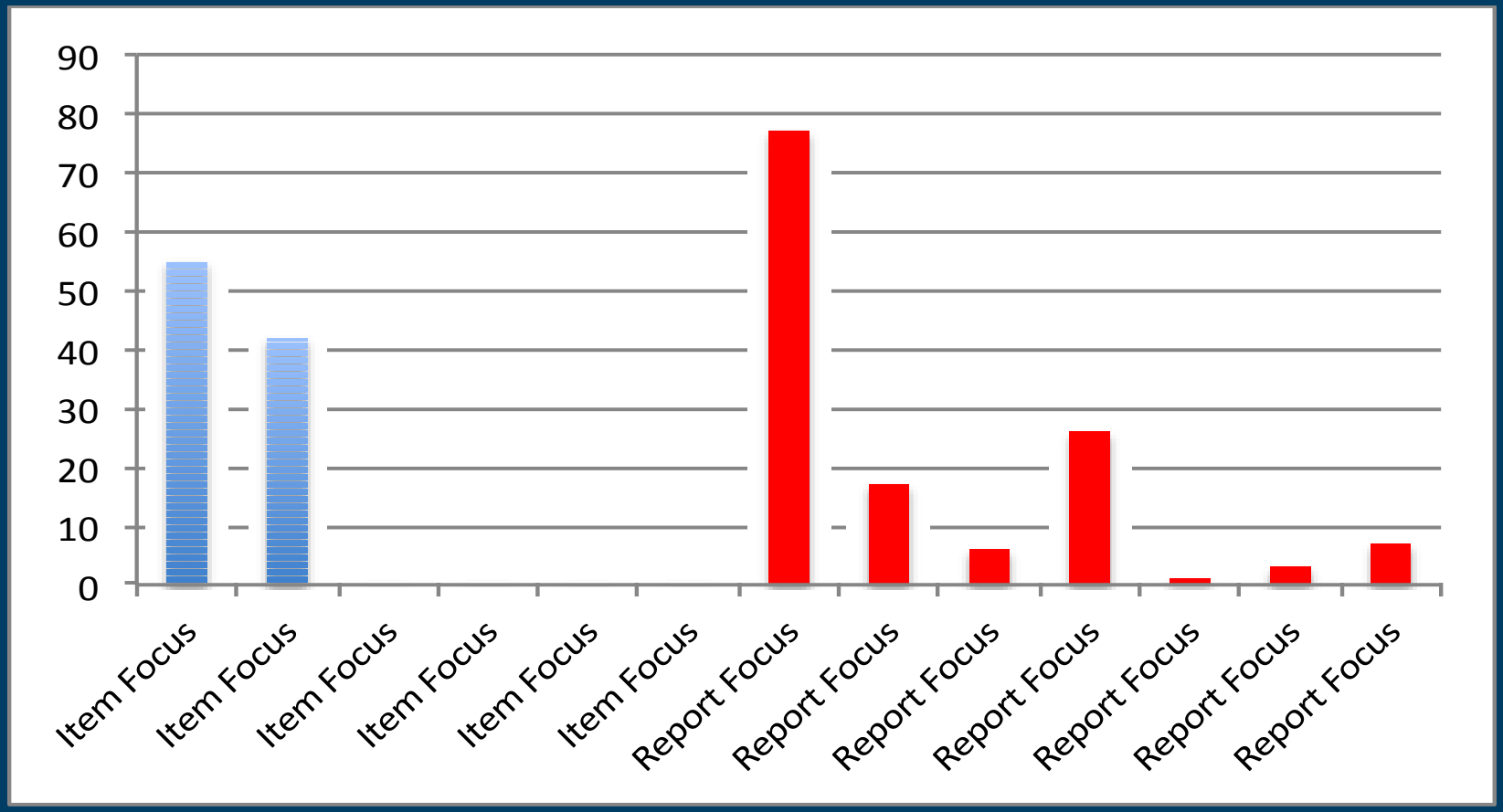
- Two different conditions emphasized different aspects of FADS.
  - The 'item development' condition
    - training focused on developing ideas for interactive assessment items, which was similar to the training session implemented during year 3.
  - The 'interpretation of reports' condition
    - Training focused on the importance of using the formative reports, exemplifying how the information in them could inform classroom practice.
- Each group had equal access to the assessments and reports

Access to the FADS reports by teachers in the two training conditions.



Note: The red bars represent access to the system by teachers in the "report-focused" training; blue bars represent the teachers that participated in the "item-focused" training.

Access to the FADS reports by individual teachers and training condition



CDMW beta trials

(a) a training session

(b) the first online assessment

(c) the second online assessment

(d) a closing interview

Online assessments

- Students of the participating teachers completed two online assessments regardless of the condition assigned to their teachers during the training session.
- The types of items included in the online assessments were drawn from the *Intermediate Constraint Taxonomy for E-Learning Assessment Questions and Tasks*

Intermediate Constraint Items*

Most Constrained → *Least* Constrained

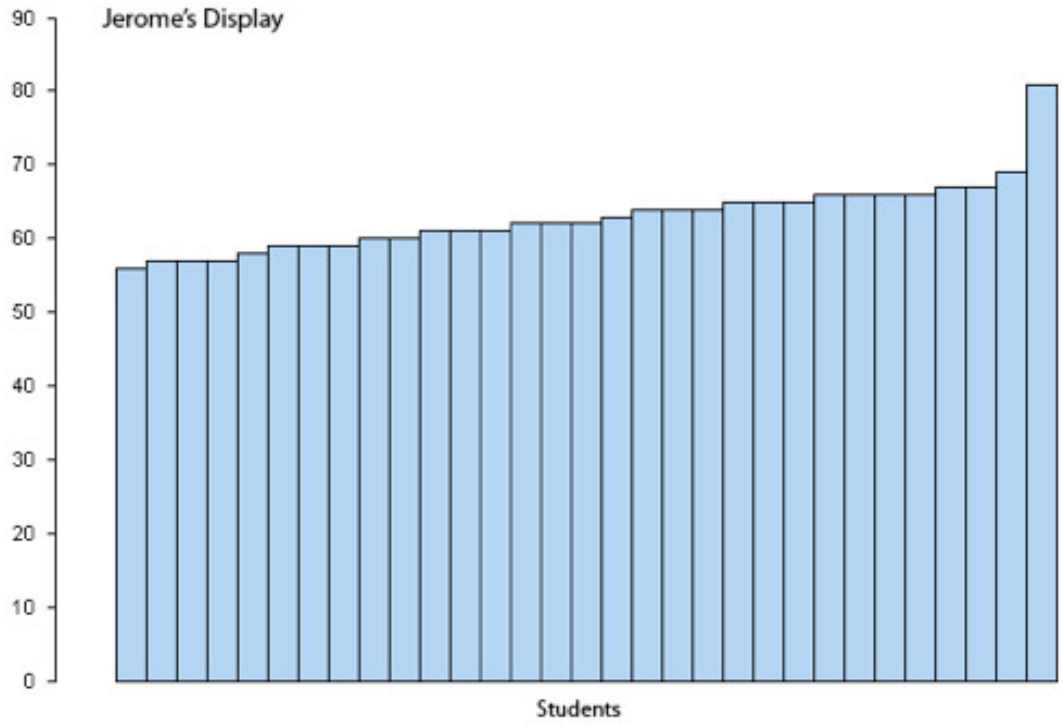
Less Complex ↓ *More* Complex

1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	7. Presentation/ Portfolio
1A. *True/False*	2A. *Multiple True/False*	3A. *Matching*	4A. *Interlinear*	5A. *Single Numerical Constructed*	6A. *Open-Ended Multiple Choice*	7A. *Project*
1B. *Alternate Choice*	2B. *Yes/No with Explanation*	3B. *Categorizing*	4B. *Sore-Finger*	5B. *Short-Answer & Sentence Completion*	6B. *Figural Constructed Response*	7B. *Demonstration, Experiment, Performance*
1C. *Conventional or Standard Multiple Choice*	2C. *Multiple Answer*	3C. *Ranking & Sequencing*	4C *Limited Figural Drawing*	5C. *Cloze-Procedure*	6C. *Concept Map*	7C. *Discussion, Interview*
1D. *Multiple Choice with New Media Distractors*	2D. *Complex Multiple Choice*	3D. *Assembling Proof*	4D. *Bug/Fault Correction*	5D. *Matrix Completion*	6D. *Essay & Automated Editing*	D. *Diagnosis, Teaching*

*Scalise, K., & Gifford, B. R. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *Journal of Teaching, Learning and Assessment, 4(6).*

examples of the developed items

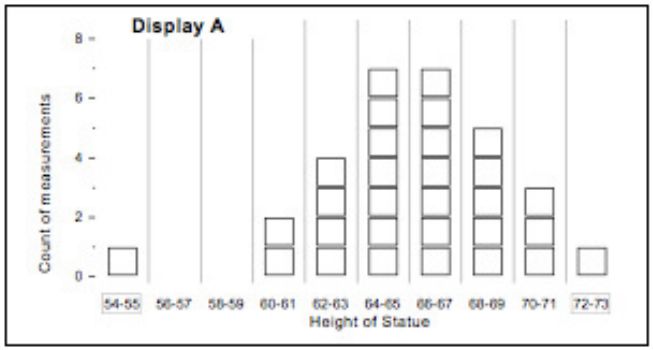
Jerome created this display of the heights of all of the students in his class:

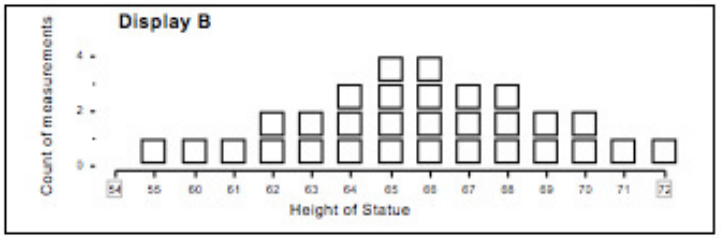


1. What does the display show about the data?
2. What does the display hide about the data?

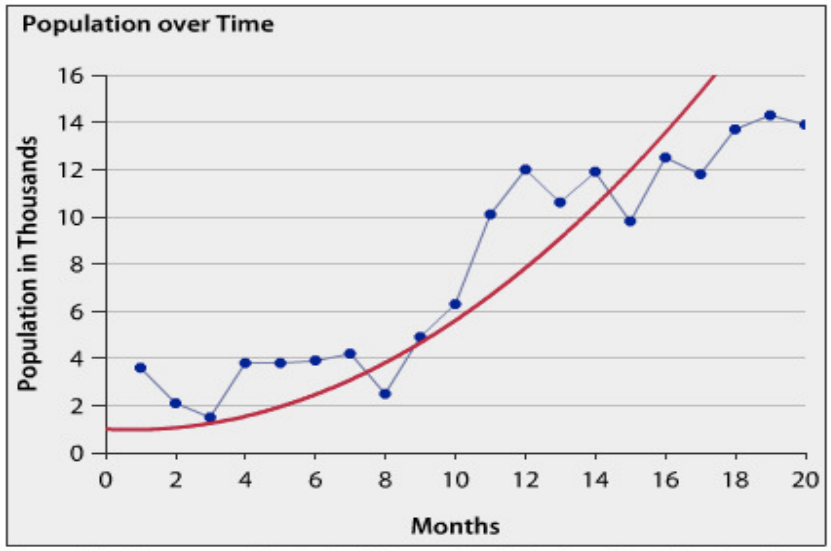
examples of the developed items

30 students measured the height of a statue in front of the city hall. Two students created displays of the data they collected. The two displays show the same data set.





1. Which display does a better job at showing how many students got each value?
   a. Display A
   b. Display B
2. Why is this display better than the other?



Custom Math Parser created by Barbara Kaskosz of the University of Rhode Island and Douglas E. Ensley of Shippensburg University.

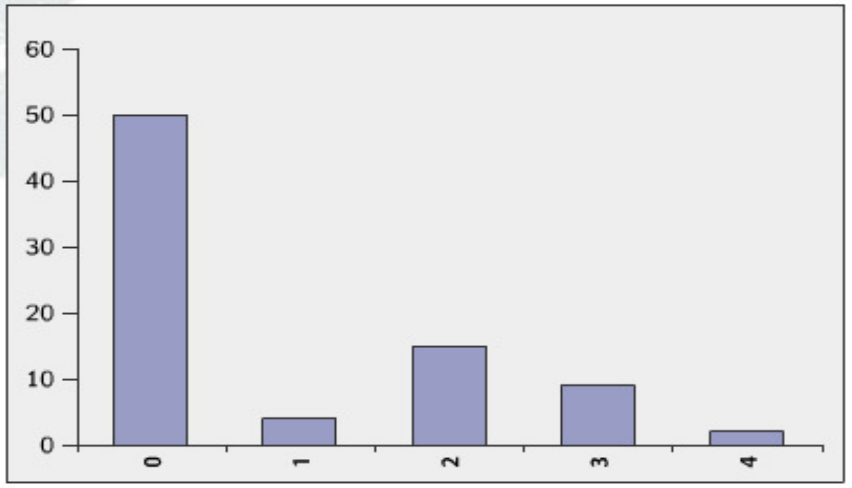
Line 2

Population

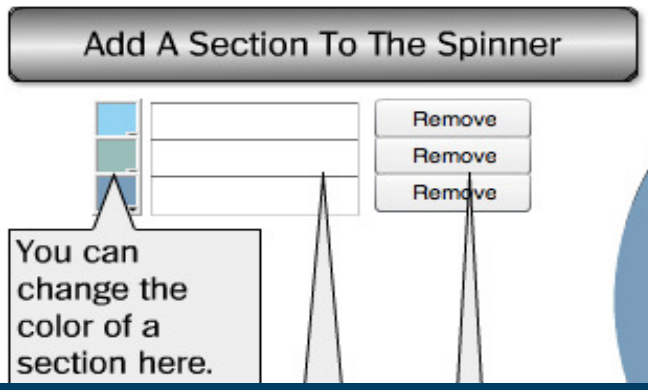
a
-1
1
0.05

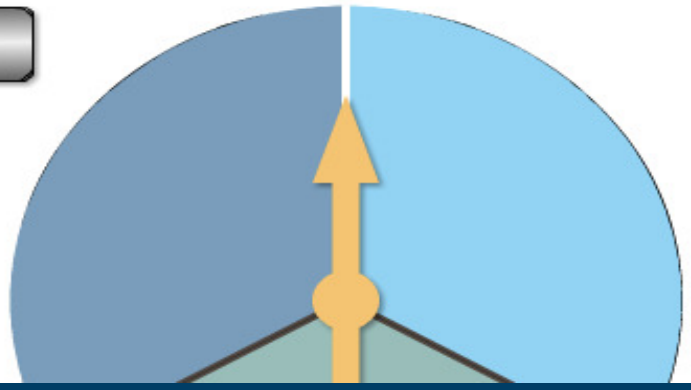
b
-10
10
-0.08

c
1



- Can you create an online "spinner" tool to help predict the chance of a female bear having a certain number of cubs? Add sections, colors and names to make your own spinner.





Online assessment 1

Activity 1

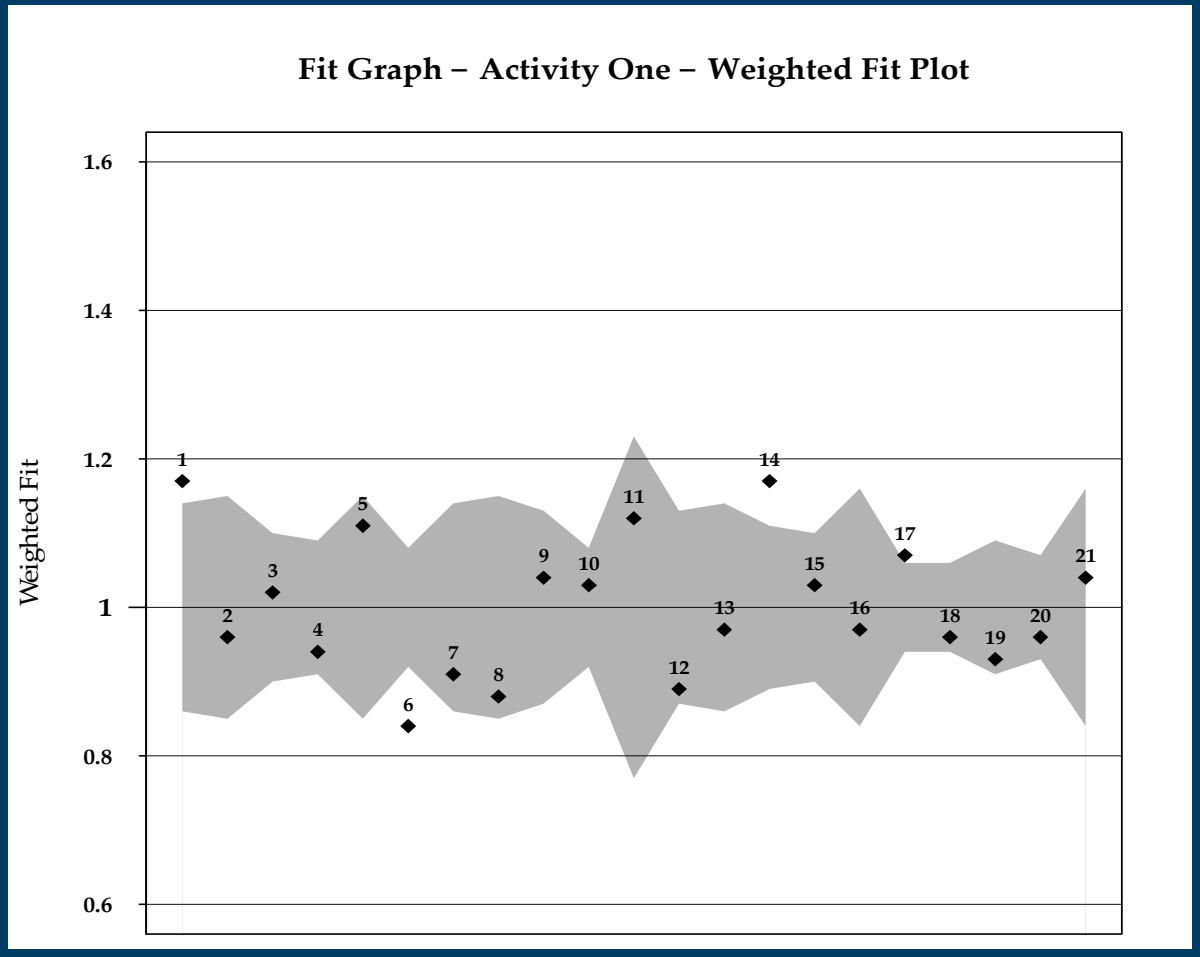
- 353 students
- 21 items.
- Partial credit model
- MLE Person separation reliability = 0.81
- EAP/PV reliability = 0.81.
- range of estimated proficiencies: -1.4 to 1.6

Online assessment 2

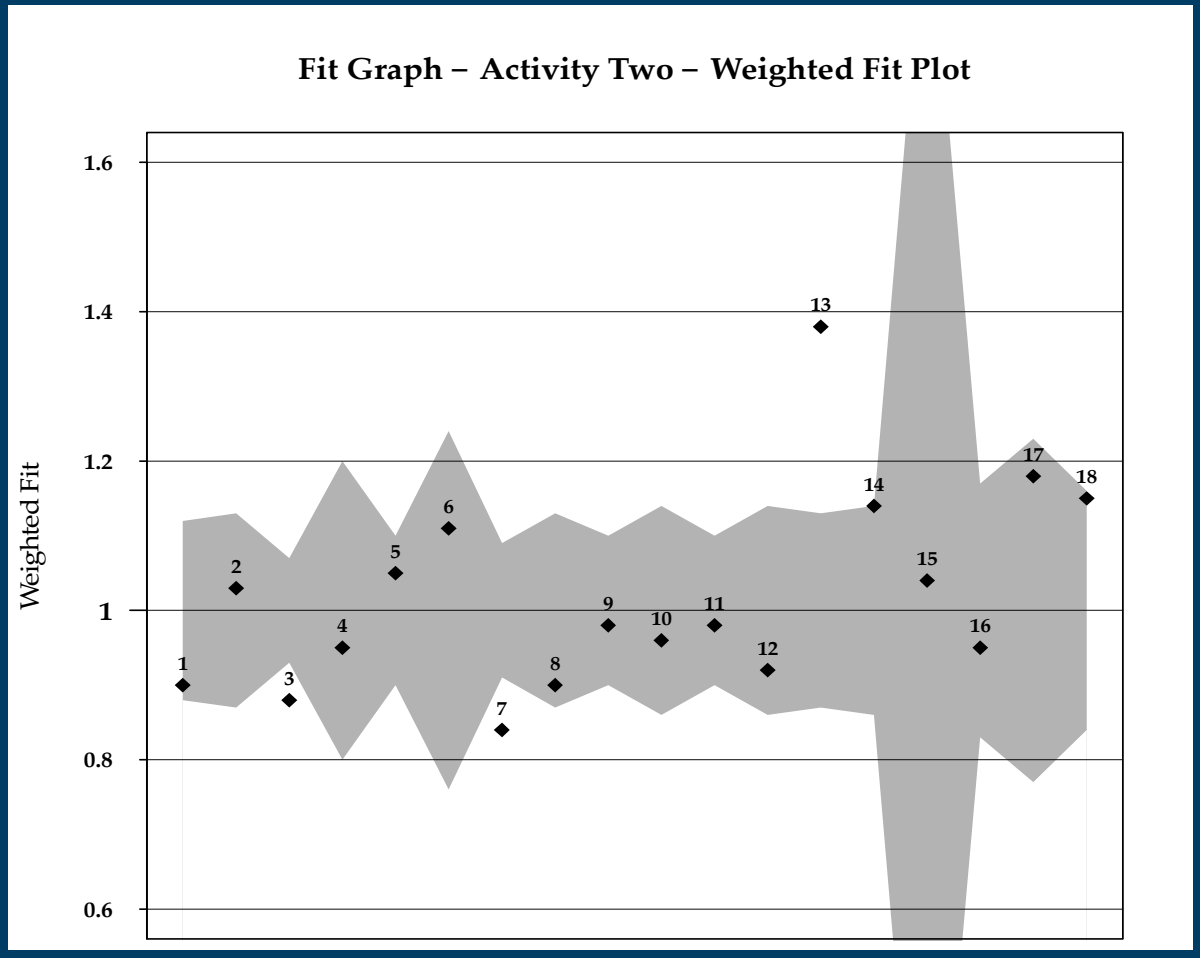
Activity 2

- 318 students
- 18 items.
- Partial credit model
- MLE Person separation reliability=0.75
- EAP/PV reliability=0.77
- range of estimated proficiencies: -1.58 to 0.92.

Activity 1 – item fit



Activity 2 – item fit



Scoring of online assessments

- Majority of task items in the trials were computer-scored
- Each online assessment included several open-ended or other questions that required hand scoring
  - These were scored in accordance with guides developed by the BEAR research team.

CDMW beta trials

(a) a training session

(b) the first online assessment

(c) the second online assessment

(d) a closing interview

Closing interview

- In-person interviews of participating teachers and students.
- Goal: to collect feedback about their experience using the FADS system.
- All interview sessions were audio-taped.

Teacher interviews

- Teachers indicated that the largest benefit of using the FADS: student assessment is scored, graded and reported automatically
  - Teachers felt these qualities saved them an immense amount of time that would otherwise be spent grading and generating statistics
  - Teachers indicated that they used the time on actual instructional planning and classroom instruction

Teacher interviews

- Teachers indicated that:
  - reports were useful and helpful in determining students' needs at both the individual and group levels
  - they used reports as a formative assessment tool to determine if any modification or review of the instruction was needed
  - they used individual level reports to identify students in the class who needed additional help, and those who did not

Teacher interviews

- Teachers indicated that students:
  - were focused and seemed generally less distracted during the FADS computerized assessments compared to paper-and-pencil assessments
  - were excited and engaged in the tasks
  - enjoyed that they had the option to move and manipulate objects (e.g. spinner wheels, pie chart segments) in some of the assessment items.

Student interviews

- Teachers were asked to invite students representative of the range of performance on the FADS assessments
- Two student interview sessions
- Each group consisted of 8-9 students.
- questions were prepared in advance

Student interviews

- Students indicated that they:
  - enjoyed the computerized assessments and compared them to colorful books.
  - were more focused during the computerized testing compared to when they took paper-and-pencil tests.
  - found erasing or inserting words from or into their answers was easier to do on computer.

Student interviews

- Students offered suggestions:
  - individual testing results on the screen at the end
  - tasks that required deeper user interaction over single click items.
  - making it possible to work in groups
  - adding instant messaging feature

Formative Assessment Delivery System

Mark Wilson, *Principal Investigator*

William P. Fisher, *Research Associate*

Samuel G. Scalise, *Co-Director, BEAR*

David Torres Irribarra, *Graduate Researcher*

Shawn Irvin, *Graduate Researcher*

Yi-Hung Lin, *Graduate Researcher*

Perman Gochyyev, *Graduate Researcher*

Kathleen Scalise, *Visiting Scholar*

Richard Vorp, *Software Engineer*

Ana Maria Albornoz, *Content Development*

Sevan Tutunciyan, *Software Engineer*

Daniel Stanfield, *Software Engineer*

Organizations

- University of California, Berkeley, BEAR Center
- Vanderbilt University
- University of Michigan

Collaborators

- Julia Koppich, *external evaluator*
- Richard Lehrer, *Vanderbilt University*
- Joe Krajcik, *University of Michigan*

Thank you

- http://bearcenter.berkeley.edu/projects/FADS/

- http://bearcenter.berkeley.edu/

Perman Gochyyev – perman@berkeley.edu
William Fisher – wfisher@berkeley.edu