

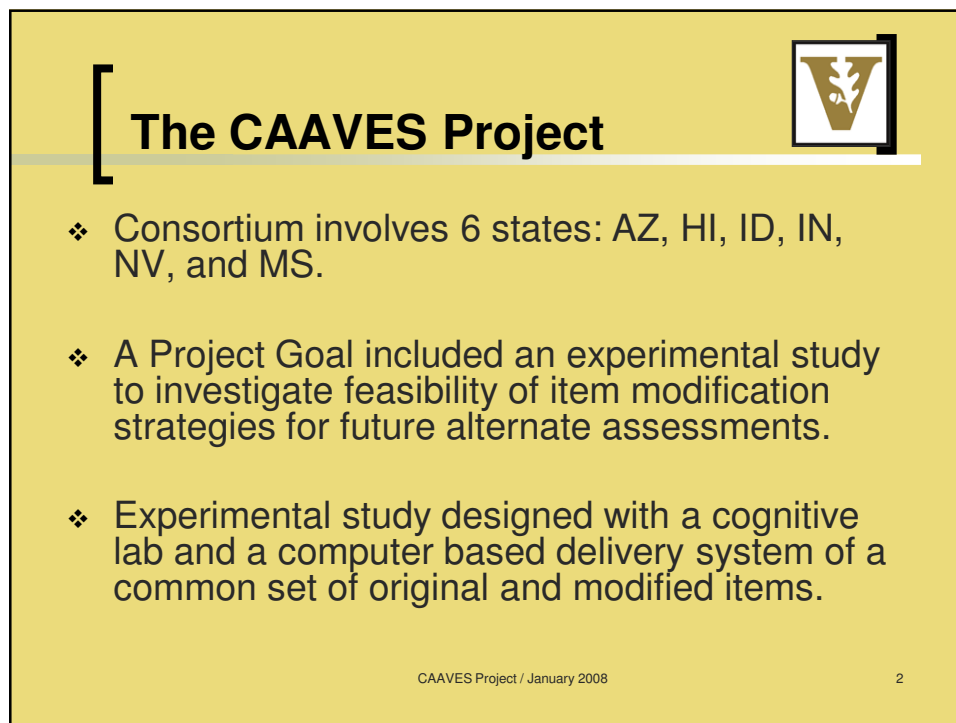
**Consortium for Alternate
Assessment Validity
and Experimental Studies**

(CAAVES Project)

Elizabeth Compton Ryan J. Kettler Andrew T. Roach

January 16, 2008

The CAAVES project is funded by the
US Department of Education, Office of Elementary and Secondary Education



The CAAVES Project

- ❖ Consortium involves 6 states: AZ, HI, ID, IN, NV, and MS.
- ❖ A Project Goal included an experimental study to investigate feasibility of item modification strategies for future alternate assessments.
- ❖ Experimental study designed with a cognitive lab and a computer based delivery system of a common set of original and modified items.

CAAVES Project / January 2008 2

The project's goal will be accomplished by...



1. Developing a common set of test items from existing reading and mathematics tests using modification principles that facilitate reading access and valid responses and
2. Using a computer-based delivery system to experimentally examine student preferences, score comparability, and item statistics of the modified items for students with and without disabilities

CAAVES Project / January 2008

3

3 Groups of Students Selected across 5 states



1. General Education Students
2. Students with Disabilities who do not meet the participation criteria set forth in the federal regulations
3. Students with Disabilities who meet the participation criteria outlined in the federal regulations

CAAVES Project / January 2008

4

[3 Criteria for the Students with Disabilities MAA Eligible group



200.1(e)(2) In the guidelines that a State establishes under paragraph (f)(1) of this section, criteria must include, but are not limited to, each of the following:

The student:

1. Has an IEP with goals based on academic content standards for the grade enrolled.

CAAVES Project / January 2008

5

[The student...



2. Has a disability precluding the student from achieving grade-level proficiency, as demonstrated by the student's performance on the state assessment or another assessment that documents academic achievement.
3. Progress to date (a) in response to appropriate instruction, is addressing the student's individual needs and (b) based on multiple measurements is such that, even if significant growth occurs, the student will not achieve grade-level proficiency within the year covered by the student's IEP.

CAAVES Project / January 2008

6

Item Modification Procedures



- **Goals of the Consortium Workgroups:**
 1. To modify existing items in reading and mathematics so they are less complex, more accessible, and likely easier yet still measure the same knowledge and skill as the original items.
 2. Document item modification procedures and guiding principles, challenges, and recommendations for future work.

CAAVES Project / January 2008

7

Started with a Menu



- ❖ Remove a response option
- ❖ Simplify language & improve readability
- ❖ Add graphic support
- ❖ Reorganize the layout

CAAVES Project / January 2008

8

Test Items from 4 Content Strands



- ❖ Reading Comprehension
- ❖ Vocabulary
- ❖ Numbers & Operations
- ❖ Data Analysis & Probability

CAAVES Project / January 2008

9

Before & After Comprehension & Vocabulary



Read the following passage and then answer questions 3 and 4.

Ready for Take Off



As I tightened my seat belt and prepared for take off, I could feel my heart pounding. Not only was it my first time flying, but I was also alone on the first leg of the journey. I was meeting Uncle Jeff in Baltimore to fly on to Boston, but the two-hour flight ahead of me felt like it would take an eternity. After the flight attendant went through the steps to take in case of an emergency, I scanned my brain to see if I had forgotten anything from home. I was pretty sure that I had not, so I settled back into my seat and tried to relax.

My heart rate and breathing rapidly increased as we lifted off the ground. I grabbed the armrests and closed my eyes. I calmed down as the pressure in the cabin stabilized and my ears began to pop. I got out my CDs and crossword puzzle book in an attempt to forget where I was.

Just when I finished my third puzzle and changed CDs, the pilot's voice echoed over the speakers, informing passengers that we were beginning our initial descent. Had it already been two hours? Was I that oblivious to the time? I discovered by looking at my watch that we had made good time and would arrive in Baltimore almost thirty minutes early.

The landing was a little bumpy but not the horrific experience I had expected. After exiting the plane, I realized how much I had dramatized such a simple experience. To my astonishment, Uncle Jeff was already waiting for me in the terminal. He had gotten there early, anticipating my arrival. When he asked how the flight had gone, what came out of my mouth next surprised even me. "Piece of cake," I replied.

3. Based on this passage, we could predict that the narrator

- A. will have less fear of flying
- B. will never fly again
- C. will always fly with friends
- D. will never fly without puzzle books

Read the following passage and then answer questions 1 and 2.

Ready for Take Off



As I tightened my seat belt and prepared for take off, I could feel my heart pounding. It was my first time flying, and I was alone. I was meeting Uncle Jeff in Baltimore, but the two-hour flight felt like it would take forever. I settled back into my seat and tried to relax.

My heart rate and breathing rapidly increased as we lifted off the ground. I grabbed the armrests and closed my eyes. I calmed down once we were in the air and I got out my CDs and crossword puzzle book.


Just as I finished my third puzzle and changed CDs, the pilot informed passengers that we were going to land soon. Had it already been two hours? Was I that **oblivious** to the time? I looked at my watch and realized we were almost there.

The landing was a little bumpy but not the awful experience I had expected. After exiting the plane, I realized how much I had worried about a simple experience. Uncle Jeff was already waiting for me in the terminal. When he asked how the flight had gone, my answer surprised even me. "**Piece of cake**," I replied.

1. Based on this passage, we could predict that the narrator

- A. will have less fear of flying
- B. will never fly again
- C. will never fly without puzzle books

10




Before & After – Numbers

4. Mr. White is buying a new car and wants to make a down payment of \$1,735.00. When he withdraws this amount from his savings account, the teller tells him that the largest bills that he can get are \$100 bills. Mr. White asks for as many hundred dollar bills as possible. How many hundred dollar bills should he receive?


A. 17
 B. 16
 C. 7
 D. 6

2. Mr. White is buying a car.

He will make a down payment of \$1,735.00.




The largest bills that he can get from the bank are \$100 bills.



At most, how many hundred dollar bills will he receive?

A. 17
 B. 16
 C. 7

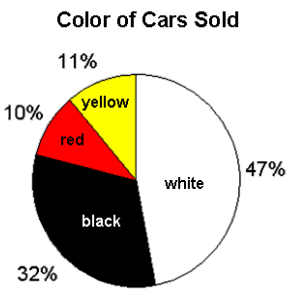
11



Before & after – Data Analysis

3. A car dealer sold 360 cars last month. The graph shows the percentage of white, black, red, and yellow cars sold.

Color of Cars Sold

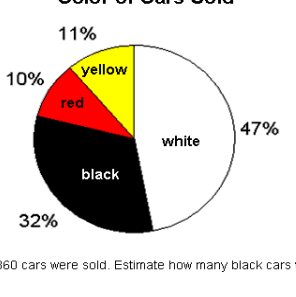


About how many were black?

A. 190
 B. 120
 C. 85
 D. 20

3. The graph shows the percentage of cars that were sold by color.


Color of Cars Sold



360 cars were sold. Estimate how many black cars were sold.

A. 190
 B. 120
 C. 85

12



Before & After – Analyze Item Statistics


8. Michael's quiz scores are 58, 86, 92, 60, and 40. What must he make on the next quiz in order to have a 70 quiz average?

A. 14
 B. 84
 C. 90
 D. 96

Criteria	Original Item																		
1. Item Difficulty (% of students getting item correct)	58%																		
2. Item Discrimination (% of students in upper group answering correct - % of students in lower group answering correct)	low- 33% high- 82%																		
3. Effectiveness of Distracters (frequency of common distracters for upper and lower groups)	B- Correct A- Least selected																		
4. Differential Item Functioning DIF																			
5. Depth of Knowledge (DOK)	2																		
6. Readability indices	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Dale-Chall</td><td>7.8</td></tr> <tr><td>Automated Readability Index</td><td>1.0</td></tr> <tr><td>Simplified Automated Readability Index</td><td>0.0</td></tr> <tr><td>FORCAST</td><td>6.7</td></tr> <tr><td>Flesch-Kincaid</td><td>3.1</td></tr> <tr><td>Rate Index (RII)</td><td>4</td></tr> <tr><td>Fry</td><td>3</td></tr> <tr><td>Raygor Estimate</td><td>3</td></tr> <tr><td>Average</td><td>3.5</td></tr> </table>	Dale-Chall	7.8	Automated Readability Index	1.0	Simplified Automated Readability Index	0.0	FORCAST	6.7	Flesch-Kincaid	3.1	Rate Index (RII)	4	Fry	3	Raygor Estimate	3	Average	3.5
Dale-Chall	7.8																		
Automated Readability Index	1.0																		
Simplified Automated Readability Index	0.0																		
FORCAST	6.7																		
Flesch-Kincaid	3.1																		
Rate Index (RII)	4																		
Fry	3																		
Raygor Estimate	3																		
Average	3.5																		
7. Accessibility rating																			
8. IRT-based Item Information	B=-.33																		
9. Frequency of Omitted Items	1.5%																		
10. Reliability of items (total scale and with individual items when one is deleted from total scale)	.828 .825 13																		

8. Mike's test scores are 58, 86, 92, 60, and 40. What score does he need on the next test in order to have a mean of 70?

A. 96
 B. 90
 C. 84




Item Accessibility & Modification Guide

- ❖ Designed by Beddow, Kettler, & Elliott for the CAAVES project.
- ❖ Purpose: to provide an organized framework for item modification.
- ❖ Closely follows guidelines from NCEO Technical Report #42. (Thompson, Johnstone, Anderson, & Miller, 2006)

CAAVES Project / January 2008


14



Theoretical Influences

- ❖ **Universal Design:** The design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. (Center for Universal Design, NC State University, 2007)
- ❖ **Cognitive Load Theory:** Learning principles that result in efficient instruction by leveraging cognitive processes. (Clark, Nguyen, & Sweller, 2006)

CAAVES Project / January 2008 15

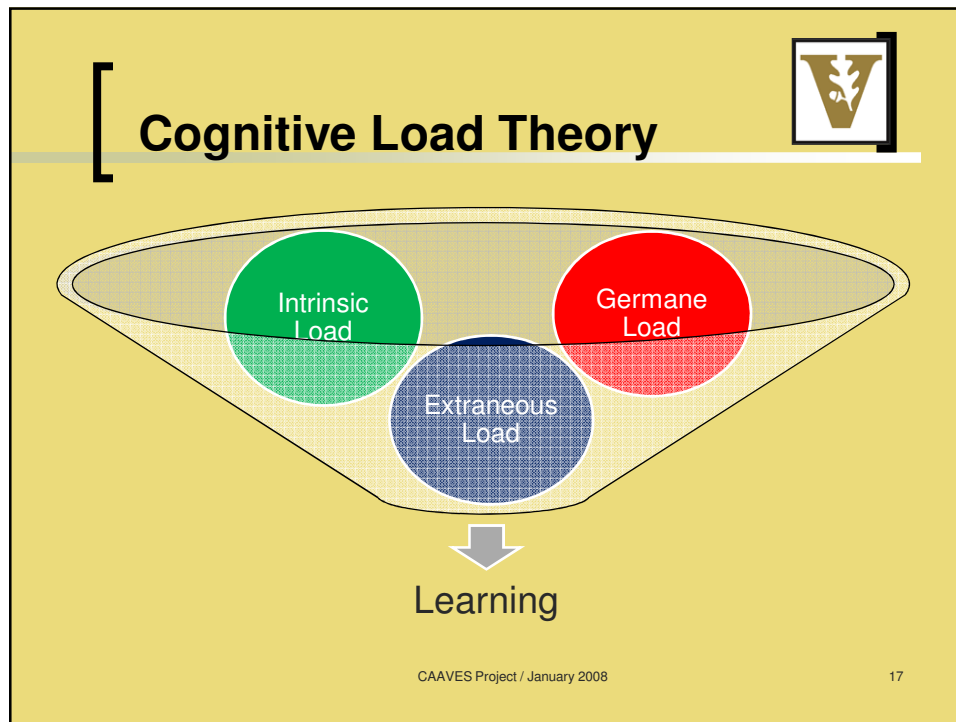


Universal Design

- ❖ Equitable use
- ❖ Flexibility in use
- ❖ Simple and intuitive
- ❖ Perceptible information
- ❖ Tolerance for error
- ❖ Low physical effort
- ❖ Size and space for approach and use

(Center for Universal Design, 2007)

CAAVES Project / January 2008 16



- ## Organization
- ❖ Two sections:
 - Considerations for All Universally Designed Assessment Items
(7 categories, 38 considerations)
 - Considerations for Computer-Based Tests
(5 categories, 26 considerations)
 - ❖ First, the rater provides a brief description of the construct the item is intended to measure.
 - ❖ Then, for each accessibility consideration, the rater is asked to check Yes, No, or N/A.
- CAAVES Project / January 2008 18

Organization



Finally, the rater is asked to circle the number that best represents his/her summative evaluation of the item's accessibility for that category based on checklist responses:

- 0 = The item is Not Accessible
- 1 = The item is Minimally Accessible
- 2 = The item is Moderately Accessible
- 3 = The item is Maximally Accessible

Categories of the *IAMG*



All Tests

- ❖ Content Relevance
- ❖ Text Content
- ❖ Text Economy/Conciseness
- ❖ Text Appearance
- ❖ Visual Content
- ❖ Fairness for Subgroups
- ❖ Format Flexibility

Computer-Based Tests

- ❖ Layout and Design
- ❖ Navigation
- ❖ Screen Reader Considerations
- ❖ Test Specific Options
- ❖ Computer Capabilities

Item Modification Cognitive Lab Overview



Our study involved 3 components:

- ❖ Students completed a series of 16 assessment items (8 reading; 8 math).
- ❖ Students were asked to think aloud as they completed or solved these items.
- ❖ We also asked follow-up questions about students' perceptions of the assessment items.

(Johnstone, Bottsford-Miller, & Thompson, 2006; Branch, 2000).

CAAVES Project / January 2008

21

Test Item Modification Distribution



	1	2	3	4	5	6	7	8
Test A	X	X	X	X				
Test B					X	X	X	X

X = Item modifications used.

CAAVES Project / January 2008

22

[Sample Size by Sub-group]

	Test A group	Test B group	Total
Students without disabilities	2	1	3
Students with disabilities; Testing Accommodations	1	2	3
Students with disabilities; Modified Alternate Assessment	1	2	3

CAAVES Project / January 2008 23

[Procedures Overview]

- ❖ We explained the think-aloud procedures, had the students restate their understanding of the process, and modeled thinking aloud on a practice item.
- ❖ We used a script adapted from a study conducted by Johnstone, Bottsford-Miller, and Thompson, 2006.
- ❖ Students were prompted only when they were silent for 10 consecutive seconds.
- ❖ If students verbalized infrequently, we reminded them to “keep thinking aloud” or “keep talking.” Otherwise we generally did not give encouragement or support.

CAAVES Project / January 2008 24

Results: Visuals and Graphs



- ❖ Most SWDs (67%) saw the visuals as being helpful and providing support on reading questions and passages.
- ❖ 100% of the students without disabilities indicated the pictures made no difference in understanding the reading questions or passages.
- ❖ Students with (50%) and without disabilities (67%) generally saw the visuals and graphs as being helpful and providing support on math items.
- ❖ ...However, 33% of SWDs indicated that the visuals/graphs were distracting or made it harder to answer the questions.

CAAVES Project / January 2008

25

Results: Directions and Bold Type



- ❖ General education students (67%) and student eligible for MAAs (100%) generally preferred test directions that were explicit, such as “Read...then answer...”
- ❖ Some students indicated that the less explicit directions (i.e., “Use the passage...”) might encourage test takers to skim rather than read closely.
- ❖ The majority of students from all groups (78% of the total) felt the use of **bold type** to identify key terms was helpful in answering the reading items.

CAAVES Project / January 2008

26

Results: 3 vs. 4 Answer Choices



- ❖ SWDs (with one exception) perceived no difference in difficulty between items having 3 or 4 possible answers on reading items.
- ❖ Conversely, 67% of the students without disabilities identified the 3-answer modification as making the reading items easier.
- ❖ The results suggest that this modification did not affect either groups' performance on most reading items [e.g., only one item ("Pesticides") had a discernable difference in student accuracy between modified and unmodified versions].

CAAVES Project / January 2008

27

Unmodified Item



Pesticides


In the late 1980s, farmers began to use a pesticide to control insects that harmed their cotton crops. This problem was solved. However, an insect group that pollinated the corn crops was also injured. Without pollination the corn kernels did not fully develop. This affected the corn harvest on which the farm families had come to depend. What is not mentioned as one effect of pesticide usage?

- A. soil contamination
- B. destruction of pests
- C. destruction of friendly insects
- D. crop losses

CAAVES Project / January 2008

28

[

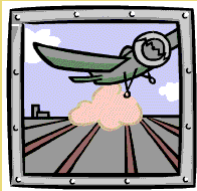


Modified Item

Pesticides
In the late 1980s, farmers began to use a chemical pesticide. It was used to control insects that harmed their cotton crops. This solved one problem, but caused another. An insect group that pollinated the corn crops was also harmed by the pesticide. Without pollination the corn kernels did not fully develop. This decreased the corn harvest.


What is **NOT** mentioned as one effect of using chemical pesticides?

- A. destruction of the soil
- B. destruction of pests
- C. destruction of friendly insects



CAAVES Project / January 200829

[



Results: 3 vs. 4 Answer Choices

- ❖ Students without disabilities (67%) and SWDs in TA group (67%) generally indicated 3 answer choices made the math items easier.
- ❖ Some students in these groups appeared to use the possible answer choices to help solve math items, but it was not clear that they used this same strategy in reading.
- ❖ For the MAA group, the 3-answer choice modification was less likely to be identified as helpful, but it did seem to make a difference on one particular item (i.e., “scientific notation”).

CAAVES Project / January 200830

Results: Analogies



- ❖ Most students (including 2/3 of SWDs) found the traditional format for the analogy easier (i.e., “meteor:space::dolphin:_____”). Some students indicated they had been taught analogies using this format and it was familiar to them.
- ❖ This was supported by the results as SWDs correctly answered all the traditional analogy items. SWDs missed items with a modified analogy format (i.e., meteor is to space as dolphin is to ____) 40% of the time.
- ❖ Familiar test item format may help make an item less difficult.

CAAVES Project / January 2008

31

Take Away Ideas



- ❖ “Conservative” modifications were used and the effects (on student performance) generally were modest. More “aggressive” modifications might result in more robust effects.
- ❖ SWDs sometimes appeared unfamiliar with concepts (e.g., percentages) or incorrectly applied problem solving strategies. In these cases, item modifications are unlikely to provide support or facilitate access.
- ❖ Reading fluency may be an issue for SWDs. Some SWDs (in the MAA group) took up to 6 minutes to read short reading passages, resulting in testing sessions that were almost twice as long as their peers. How could (or should) technology be used to address this barrier?

CAAVES Project / January 2008

32

References



- ❖ Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think-alouds and think-afters. *Library and Information Science Research*, 22(4), 371–392.
- ❖ Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Tech44/>

CAAVES Project / January 2008

33

Experimental Study Progress



- ❖ November 2007- Consortium members reviewed Cognitive Lab Results
- ❖ Lessons Learned to date:
 - ❖ Modifications appear too conservative
 - ❖ Reflect & consider Cognitive Load Theory
 - ❖ Participation criteria & identification process takes time
 - ❖ Involve general education content specialists, special education specialists, and measurement/assessment specialists in the item modification process

CAAVES Project / January 2008

34

Experimental Study Progress



- ❖ New test design: 13 original, 13 modified without read aloud, 13 modified with read aloud
- ❖ Random assignment of students to a sequence of the items
- ❖ Read aloud feature will have a voice read directions and question stems. However, the reading items may or may not use this feature for the question stem.
- ❖ Assessment window February 2008
- ❖ Preliminary results – June 2008 at National Assessment Conference in Orlando

CAAVES Project / January 2008

35

Any Questions?



- ❖ Elizabeth Compton, Idaho Department of Education & Boise State University, ebermen@boisestate.edu
- ❖ Ryan J. Kettler, Vanderbilt University, ryan.j.kettler@vanderbilt.edu
- ❖ Andrew T. Roach, Georgia State University, aroach@gsu.edu
- ❖ CAAVES Website: www.vanderbilt.edu/caaves/html

CAAVES Project / January 2008

36