SELECTING STATE ASSESSMENT ITEMS FOR INSTRUCTIONAL SENSITIVITY

John Fremer
Caveon Test Security

Instructional Sensitivity Conference
November 2013
H. L. MENCKEN

• For every complex problem, there is an answer that is:

  – Clear
  – Simple
  – Wrong
TWO ITEM PRETEST

• #1  - What is a Foxaluksen?

• #2 - What is the name of the first button on a shirt, blouse, jacket?

• Please raise your hand if you know the answer to:
  – Question # 1
  – Question # 2
ANSWER TO QUESTION - #1

• A Foxaluksen =
  – Small animal with both feathers and fur
  – Yellow
  – Lives underground in suburban San Jose, CA
  – Shy, seldom seen

• Please consult with the people next to you to be sure you now know what a Foxaluksen is.
• Buttons have names
  – This fact is little known outside the Fremer family
  – It does not matter what the garment is
  – The first button is called “Foozel”

• Please consult with your neighbors to be sure you all know the name of the first button.
POST TEST

• #1 - What is a Foxaluksen?

• #2 - What is the name of the first button on a shirt, blouse, jacket?

• Please raise your hand if you know the answer to:
  – Question # 1
  – Question # 2
INSTRUCTIONAL SENSITIVITY OF QUESTIONS

- **Pretest Scores**
  - Question # 1 = Zero
  - Question # 2 = Zero

- **Posttest Scores**
  - Question # 1 = 100% Correct
  - Question # 2 – 100% Correct

- Instructional Sensitivity = Stunning
OUTLINE OF REMAINING PRESENTATION

1. MAIN POINT
2. AREAS OF AGREEMENT
3. MY PERSPECTIVE
4. CONCEPTUAL MODEL OF ACCOUNTABILITY TESTING
5. KEY ASSUMPTION
6. USING EVIDENCE OF LOW PERFORMANCE
7. A PERSONAL EXAMPLE
1. MAIN POINT

- Instructional sensitivity is a wrong-headed way to select items for educational accountability tests
2. MY PERSPECTIVE

• Main Line - Testing Industry and Profession
  – 35 Years ETS
  – 4 years Company Now Part of Pearson
  – 10 years Caveon Test Security

• My Fellowship to Graduate School Awarded by Robert L. Thorndike

• 50 Million People have Taken Tests I Had a Role in Developing
3. AREAS OF AGREEMENT

- We admire and respect teachers
  - My wife of 54 years was a teacher
  - Our youngest daughter has just moved to a “mentor” role from being a 5th grade teacher in the San Jose Schools
- We want educational accountability testing to play a constructive role
- We believe that testing influences instruction (because we are not fools)
4. MODEL OF ACCOUNTABILITY TESTING

• Standards Carefully Developed
• Questions Written to Test Standards
• Tests Administered, Scored, Reported
• Good Performance is Celebrated
• Areas of Low Performance Receive Additional Attention
5. KEY ASSUMPTION

• Students are Likely to Learn What we Focus on (aka WYTIWYG)

• If Students are not Learning Some Objectives to our Satisfaction
  – We should focus on those areas
  – We should try different approaches

• If we do a Better Job of Teaching these Objectives Student Performance will Improve
5. KEY ASSUMPTION (cont.)

• What if Students Still do Poorly?
  – Our expectations may be unrealistic
  – May be improper placement of objective for grade level or sequence of instruction
6. USING EVIDENCE OF LOW PERFORMANCE

• **Not** to Throw Out Items that Test Important Standards

• Re-examine Standards, Objectives, and Items

• Is the Problem, for example, a Misconception?
6. USING EVIDENCE OF LOW PERFORMANCE (cont.)

- Does the Test Design Support Generalizations across Items that could Help Teachers?

- If we Use the Available Information Wisely, Performance will Improve
7. ANOTHER PERSONAL EXAMPLE

• ETS Workshop on Criterion Referenced Testing
  – Michael Zieky and I were the leaders
  – We used pre and post testing of workshop content
  – Goal was to show the value of this kind of data
7. A PERSONAL EXAMPLE (cont.)

• We Ran Item Analyses of Both Sets of Tests

• We Looked for Items with
  – High increases in performance
  – Low increases in performance

• You Could Call this an “Instructional Sensitivity” Analysis
7. A PERSONAL EXAMPLE (cont.)

• What did We Find?

• High growth – very specific to course, e.g., “The ETS item difficulty index is known as [Answer - Delta]

• Items that it Would be Hard to Learn Outside of the Course

• Items Very Susceptible to “Memory Drills”
7. A PERSONAL EXAMPLE (cont.)

• Low Growth – items requiring understanding of measurement concepts, e.g., “Why not pick from your item pool only the items most highly correlated with the total score when constructing a test?”

• Answer – if you do so you will limit the coverage of your test and it will be a less valid indicator of what you are trying to measure.
WHAT DID I SAY?

• Instructional Sensitivity is not an appropriate basis for choosing items for educational accountability tests

• Reduces the critical role of standards

• Distracts from the educationally sound use of test data to help focus instruction

• Could be a slight but unwelcome push towards a focus on “drill and practice” content and we suffer from that already