Assessing NSF Programming--Standards-Based Reform
Assessment Technologies: The San Francisco Project

Richard J. Shavelson and Maria Araceli Ruiz-Primo

Stanford University

Performance Effectiveness Review
National Science Foundation

January 27, 1999
Achievement Indicators: Instructional Sensitivity

Depth of Assessment Probe

Remote: Standardized National Science Achievement Tests

Distal: Large-Scale Performance Assessment from State/National Curriculum Framework

Proximal: Same Concept/Principle--New Investigation

Close: “Embedded” Assessments -- A Slightly More Advanced Activity in Unit

Immediate: Lab Notebooks & Classroom Tests

Classroom Instruction
Agenda

PART I: Framework for Evaluating Science Education Reform
• A Working (and Evolving) Definition of Science Achievement
• Linking Assessments to Components of Achievement
• Multilevel Achievement Assessment

PART II: The San Francisco Study
• The Proximity of the Assessments
• The Study
• The Findings

PART III: Concluding Comments
• What We Have Learned
• The Larger Picture
Toward An Achievement Framework: Knowledge Components

- **Declarative Knowledge** (Knowing the “that”)
  - Domain-specific content:
    - facts
    - concepts
    - principles

- **Procedural Knowledge** (Knowing the “how”)
  - Production system--condition-action rules

- **Strategic Knowledge** (Knowing the “which,” “when,” and “why”)
  - Problem schemata/
    - strategies/
    - operation systems

- **Characteristics That Vary According to Proficiency Level**
  - Extent (How much?)
  - Structure (How is it organized?)
  - Others (Precision? Efficiency? Automaticity?)

- **Cognitive Tools:**
  - Planning
  - Monitoring
### Linking Assessments to Achievement Components

<table>
<thead>
<tr>
<th>Extent</th>
<th>Declarative Knowledge</th>
<th>Procedural Knowledge</th>
<th>Strategic Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-Choice</td>
<td>Performance Assessments</td>
<td>Performance Assessments</td>
<td></td>
</tr>
<tr>
<td>Conceptual Maps</td>
<td>Procedural Maps</td>
<td>Models/Mental Maps</td>
<td></td>
</tr>
</tbody>
</table>

**Others**

- Performance Assessments
- Models/Mental Maps
Achievement Indicators: Instructional Sensitivity

Depth of Assessment Probe

Remote: Standardized National Science Achievement Tests

Distal: Large-Scale Performance Assessment from State/National Curriculum Framework

Proximal: Same Concept/Principle--New Investigation

Close: “Embedded” Assessments -- A Slightly More Advanced Activity in Unit

Immediate: Lab Notebooks & Classroom Tests

Classroom Instruction
Variables Unit: Swingers Activity

**Goal:** Gain experience with the concepts of system, variable, and controlling and manipulating variables

**Assessment Task:** Same concept and (high) structure, slightly different materials and measurement method

**Unit Activity**

**Close Assessment Task**
**Variables Unit: Lifeboats Activity**

**Goal:** Gain experience with the concepts of system, variable, and controlling and manipulating variables

**Assessment Task:** Same concept, different structure (low), materials, and measurement method

---

**Unit Activity**

**Proximal Assessment Task**
Mixtures & Solutions Unit: Reaching Saturation Activity

**Goal:** Gain experience with the concepts of mixtures and solutions, concentration, saturation, and chemical reaction

**Assessment Task:** Same concept, slightly different structure and materials and different measurement method

**Unit Activity**

**Close Assessment Task**
Mixtures & Solutions Unit: Fizz Quiz Activity

Goal: Gain experience with the concepts of mixtures and solutions, concentration, saturation, and chemical reaction

Assessment Task: Same concept, slightly different materials, different (low) structure and measurement method

Unit Activity

Proximal Assessment Task
CSIAC Performance Assessment: Is All the Trash the Same?

Distal Assessment Task

**TASK 1**

**Sort your trash**
- Take all of the trash items out of Bag A and put them on your placemat.
- Observe each trash item.
- Sort the trash items into groups based on characteristics the items have in common.
- Using all of the trash items, make at least three groups. Do not make more than five groups.
- Each trash item can be in only one group.
- Draw a circle around each group on your placemat.

**TASK 2**

**Identify your trash**
- Look at the "Trash Identification Table" below.
- Find a picture of each trash item that is on your placemat.
- Find the number that matches each trash item in the "Trash Identification Table."
- Write the number for each trash item on your placemat next to each item.
- Make sure that you put a number by every trash item.

**TASK 3**

**Trash Chart**
- Record the information from your placemat onto the chart below. Write your reasons for putting the trash in each group under "Reasons."

---

**Trash Identification Table**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Nail</td>
<td>Cupcake Liner</td>
<td>Newspaper</td>
<td>Construction Paper</td>
<td>Peanuts</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Paper Clip</td>
<td>Bean</td>
<td>Aluminum Foil</td>
<td>Wire</td>
<td>Styrofoam</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>String</td>
<td>Macaroni</td>
<td>Staples</td>
<td>Twist Tie</td>
<td>Plastic</td>
</tr>
</tbody>
</table>
# Proximity Profile of the Assessments

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>MIXTURES &amp; SOLUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proximal</td>
</tr>
<tr>
<td>Close</td>
<td></td>
</tr>
<tr>
<td>Proximal</td>
<td>Pendulum</td>
</tr>
<tr>
<td>Distal</td>
<td>Bottles</td>
</tr>
</tbody>
</table>

- **Purpose**: Close, Proximal, Distal
- **Content**: Close, Proximal, Distal
- **Task**: Close, Proximal, Distal
- **Directedness**: Close, Proximal, Distal
- **Materials**: Close, Proximal, Distal
- **Measurement Methods**: Close, Proximal, Distal

**Proximal Profile**
- **VARIABLES**
  - Purpose: Close, Proximal, Distal
  - Content: Close, Proximal, Distal
  - Task: Close, Proximal, Distal
  - Directedness: Close, Proximal, Distal
  - Materials: Close, Proximal, Distal
  - Measurement Methods: Close, Proximal, Distal

**Distal Profile**
- **VARIABLES**
  - Purpose: Close, Proximal, Distal
  - Content: Close, Proximal, Distal
  - Task: Close, Proximal, Distal
  - Directedness: Close, Proximal, Distal
  - Materials: Close, Proximal, Distal
  - Measurement Methods: Close, Proximal, Distal

**Close Profile**
- **VARIABLES**
  - Purpose: Close, Proximal, Distal
  - Content: Close, Proximal, Distal
  - Task: Close, Proximal, Distal
  - Directedness: Close, Proximal, Distal
  - Materials: Close, Proximal, Distal
  - Measurement Methods: Close, Proximal, Distal
Study Questions

- Does hands-on science instruction impact students’ performance?
- If so, does the magnitude of impact differ depending on the proximity of the assessment to the curriculum?
- Are findings replicable across curricular units?
Study Design

Participants: Ten schools from a medium sized urban school district in the Bay Area. Twenty teachers and about 500 fifth-graders.

Design: Replications across classrooms: Initial Student Status, Unit Implementation, Final Student Status

Instrumentation: FOSS UNITS

<table>
<thead>
<tr>
<th>Variables (Fall)</th>
<th>Mixtures and Solutions (Spring)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td>Proximal</td>
</tr>
<tr>
<td>Pendulum</td>
<td>Bottles</td>
</tr>
<tr>
<td>Saturation</td>
<td>Mystery Powders</td>
</tr>
<tr>
<td></td>
<td>Distal</td>
</tr>
<tr>
<td></td>
<td>CSIAC-PA &amp; MC</td>
</tr>
</tbody>
</table>
## Within-Classroom Design

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Initial Status Pretest</th>
<th>Implementation</th>
<th>Final Status Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Close</td>
<td>Science Journals</td>
<td>Close CSIAC-PA &amp; MC</td>
</tr>
<tr>
<td>2</td>
<td>Proximal</td>
<td>Science Journals</td>
<td>Proximal CSIAC-PA &amp; MC</td>
</tr>
</tbody>
</table>
Instructional Sensitivity: Preliminary Results

[Bar chart showing mean gain in SD units for different items: Close and Proximal. Items include Solutions, Pendulum, Bottles, and Mystery Powders. Solutions have the highest mean gain.]
Close Assessments: Effects Across Classrooms

![Bar chart showing mean gain in SD units across classrooms for Pendulum and Solutions.]
Proximal Assessment: Effects Across Classrooms

![Bar Chart]

- Mean Gain in SD Units
- Bottles
- Mystery Powders

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1 0.8 0.6 0.4 0.2 0 0.2 0.4 0.6 0.8 1
## Correlations Among Assessments
(Pilot Study)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Time</th>
<th>Variables</th>
<th>Mixtures &amp; Solutions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Directedness of Assessment</td>
<td>Posttest</td>
<td>CSIAC PA</td>
<td>Directedness of Assessment</td>
</tr>
<tr>
<td>1</td>
<td>Pretest</td>
<td>Structured</td>
<td>.76</td>
<td>.64</td>
<td>Less-Structured</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>Structured</td>
<td>.64</td>
<td></td>
<td>Less-Structured</td>
</tr>
<tr>
<td>2</td>
<td>Pretest</td>
<td>Less-Structured</td>
<td>.71</td>
<td>.03</td>
<td>Structured</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>Less-Structured</td>
<td>.20</td>
<td></td>
<td>Structured</td>
</tr>
</tbody>
</table>
### Immediate Assessment: Students’ Science Journals

#### Unit Implementation
- What instructional activities were implemented as reflected in the students’ journals
- Were other appropriate additional activities implemented?

#### Students Performance
- Were students communications complete, focused, organized?
- Did students’ communications indicate conceptual and procedural understanding?

#### Teacher’s Feedback
- Did the teacher provide helpful feedback on students’ performance?
- Did the teacher encourage students to reflect on their work?
Example of a Journal

Implementation of an Instructional Activity
Variation in Implementation Across Classrooms: Mixtures & Solutions (Pilot Study)

Unit Implementation (Max 24)

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
</tr>
</tbody>
</table>

“Close” Effect Size

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean Gain in SD Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>1.2</td>
</tr>
</tbody>
</table>
What We Have Learned

• Instruction had an impact on students’ performance
• Results were on the predicted direction across units: Close assessments were more sensitive to changes in students performance than Proximal assessments
• Sensitivity of Distal assessments cannot be evaluated since pretest data were not available
• High variation in the quality of implementation of hands-on science instruction
• High between-class variation in effect sizes
• The use of more distal or remote measures may lead to an erroneous conclusion that the reform has no impact
• However, if impact is only evident at the closest level, this raises questions about the reform itself
The Larger Picture

- More than one source of evidence should be used to evaluate the impact of instruction on students’ achievement. Evidence should bear on declarative, procedural, and strategic knowledge.

  Warning: Attention to different types of knowledge (and corresponding assessments) may lead to changes in curriculum!

- E&HR should support
  - Curriculum developers’ efforts to put assessments into their curricula
  - One or more Centers that serve as a resource to educators as they attempt to build and implement assessment in teaching
  - Basic research into the quality of alternative assessments

- E&HR should make assessment research a focus