Alignment in Educational Testing: What it is, What it isn’t, and Why it is Important

Stephen G. Sireci

University of Massachusetts Amherst

Presentation delivered at the Connecticut Assessment Forum

Rocky Hill, CT, August 13, 2010

(c) Stephen G. Sireci, 2010
What is Alignment?

- To arrange in a line so as to be parallel
- The proper positioning or state of adjustment of parts
Aligning Curriculum, Instruction, & Assessment (to facilitate student learning)

- Curriculum
- Instruction
- Assessment

(c) Stephen G. Sireci, 2010
Align What to What?

- Instruction to Curriculum?
- Assessment to Curriculum?
- Assessment to Instruction?

And,

to what instruction?
to what curriculum?
Take-home message #1

- Educators have different reasons for evaluating alignment.
- Thus, the goals of an alignment study should be clearly specified in advance (before deciding on alignment method).
In this presentation,

- I will discuss issues & methods in evaluating curriculum/assessment alignment
- Relate alignment research to content validity research
- Provide suggestions for evaluating test/curriculum alignment
Facilitating Student Learning through Alignment

Webb (1997): Alignment is the “Degree to which expectations and assessments are in agreement...and guide the system towards students learning what they are expected to know and do” (p. 4).
La Marca et al. (2000)

“Alignment is . . . the degree to which assessments yield results that provide accurate information about student performance regarding academic content standards at the desired level of detail, to meet the purposes of the assessment system” (p. 24).

(c) Stephen G. Sireci, 2010
Validity and Alignment

What is validity?

“Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 11)
And so evidence of “alignment”

- is primarily validity evidence based on test content.

**Standards** 5 sources of validity evidence:

1. Test content
2. Internal structure
3. Relations to other variables
4. Response processes
5. Testing consequences
Content validity and “alignment”

- Alignment is a newer concept, similar to content validity, that deals with the degree to which tests and test specifications are congruent with curriculum frameworks and assessment goals.

- Most, but not all, elements of alignment can be subsumed under content validity.
What is “content validity?”

The degree to which the content of a test is congruent with the purposes of the testing.

4 Elements of CV:
- Domain definition
- Domain relevance
- Domain representation
- Appropriate test construction procedures

Sireci (1998a,b, 2009)
Methods for evaluating content validity/alignment

Methods involve:

- Subject matter experts (SMEs)
- Reviewing test items
- Gathering judgmental data
- Summarizing the data
Examples of traditional content validity rating tasks

(a) “Match each test item to the content area you believe it measures…”

or

(b) “Read objective and rate the degree to which each item measures it.”
(a) Please match each item to 1 of the 3 science content areas:

<table>
<thead>
<tr>
<th>Item</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(b) Rate the congruence of each item to each objective where 1=high congruence, 0=medium congruence, and –1=no congruence:

<table>
<thead>
<tr>
<th>Objective: Use math to solve real world problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
Alignment studies

- Also use subject matter experts, but have extended the rating tasks to focus not only on the assessments and test specifications, but also on the curriculum or instruction
What do we evaluate in an alignment study?

“assessments must adequately cover the content standards with appropriate depth, reflect [their] emphasis, provide scores that cover the range of performance standards, allow all students an opportunity to demonstrate their proficiency, and be reported in a manner that clearly conveys student proficiency as it relates to the content standards” (La Marca, et al., 2000, p. 24).

(c) Stephen G. Sireci, 2010
Evaluating Test Content via “Alignment”

- Specifically referenced in NCLB
- Many different “models” or methods
  - Webb
  - La Marca
  - Porter (Surveys of Enacted Curriculum)
  - Achieve
  - Hybrids

(see Bhola, Impara, & Buckendahl, 2003; Martone & Sireci, 2009)
Webb Methodology: SMEs’ tasks

- Categorical concurrence: match test items to framework standards
- Depth-of-knowledge consistency:
  - rate cognitive complexity of objectives w/in standards and of items measuring them
- Range-of-knowledge consistency:
  - # of benchmarks w/in standard measured by > 1 item
- Balance of representation
  - how evenly distributed are items across objectives?

(c) Stephen G. Sireci, 2010
Achieve Methodology

- **Individual item focus**
  - Confirm test blueprint
  - Content centrality – degree of content match (2, 1A, 1B, 0)
  - Performance centrality – cognitive demand match (2, 1A, 1B, 0)
  - Source of challenge – fairly constructed items (1, 0)
  - Level of cognitive demand of item
Achieve Methodology (2)

- **Sets of items (or entire test)**
  - Level of challenge – degree of overall challenge (qualitative)
  - Balance – degree the items match the overall expectations of the standard (qualitative)
  - Range – degree have overall standard coverage (fraction of objectives measured)
SEC Methodology

- Common content language
- Degree of alignment
  - Each cell in 2-D matrices (content by expectations) represents proportion of content, assessment, or standards in that cell
  - Alignment index – sum of cell-by-cell intersects (0-1)
  - Topographical map layouts – relative areas of concentration and easier comparisons
Online resources:
www.seconline.org

The following 2 slides are from www.seconline.org and illustrate how teachers fill out a survey.
Reporting Instructional Content

Step 1: Report time spent on topics taught

Review the list of topics presented for the current Content Area. For each topic in the list that is taught to the target class, select a radio button corresponding to 1, 2, or 3 based on the following definitions:

0 = Not covered
1 = Less than 1 lesson
2 = 1-5 lessons
3 = more than 5 lessons

<table>
<thead>
<tr>
<th>Topic</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole numbers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decimals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio, proportion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patterns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real numbers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponents, scientific notation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors, multiples, divisibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds, evens, primes, composites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order of operations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationships between operations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical properties (eg. distributive properties)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continue | Next
Reporting Instructional Content

Step 2: Report expectations for students for each topic taught.

For each topic selected from the previous screen set the cognitive expectations for students for each of 5 categories of cognitive demand, using the following definitions:

0 = No emphasis
1 = Slight emphasis
2 = Moderate emphasis
3 = Sustained emphasis

<table>
<thead>
<tr>
<th>Expectations for Students in Mathematics</th>
<th>Memorize Facts, Definitions, Formulas</th>
<th>Perform Procedures</th>
<th>Demonstrate Understanding of Mathematics Concepts</th>
<th>Generalize, Conjecture, Prove</th>
<th>Solve non-routine problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place Value</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Whole Numbers</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Operations</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Fractions</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Decimals</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Percents</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
<tr>
<td>Ratio, proportion</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
<td>0-None</td>
</tr>
</tbody>
</table>

Continue Previous
The following 2 slides are from the Ohio State Dept. of Ed

- They illustrate some results from an SEC study.
Summary

- **Webb**: detailed quantified results
- **Achieve**: evaluates blueprint, captures item and objective quality, consensus among reviewers, qualitative overview, detailed reports
- **SEC**: common framework of topics, evaluates instructional alignment, less on alignment quality, helpful for professional development

(c) Stephen G. Sireci, 2010
### Summary: 3 Most Popular Alignment Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Dimension</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Webb (1997)</strong></td>
<td><em>Categorical Concurrence</em></td>
<td>Match of items to general content areas</td>
</tr>
<tr>
<td></td>
<td><strong>Depth of Knowledge Consistency</strong></td>
<td>Cognitive level of items compared to cognitive level of benchmark/objective</td>
</tr>
<tr>
<td></td>
<td><strong>Range of Knowledge Correspondence</strong></td>
<td>Number of benchmarks/objectives measured within general content area</td>
</tr>
<tr>
<td></td>
<td><strong>Balance of Representation</strong></td>
<td>Distribution of items across general content areas</td>
</tr>
<tr>
<td><strong>Achieve (2006)</strong></td>
<td><em>Content Centrality</em></td>
<td>Congruence between item &amp; objective/benchmark</td>
</tr>
<tr>
<td></td>
<td><em>Performance Centrality</em></td>
<td>Congruence between cognitive demand of item and objective/benchmark</td>
</tr>
<tr>
<td></td>
<td><strong>Source of Challenge</strong></td>
<td>Grade-level appropriateness</td>
</tr>
<tr>
<td></td>
<td><em>Level of Cognitive Demand</em></td>
<td>Cognitive level measured by item</td>
</tr>
<tr>
<td></td>
<td><strong>Level of Challenge</strong></td>
<td>Degree to which test captures difficulty implied by general content areas</td>
</tr>
<tr>
<td></td>
<td><strong>Balance</strong></td>
<td>Holistic evaluation of how well test represents content/cognitive specs</td>
</tr>
<tr>
<td></td>
<td><strong>Range</strong></td>
<td>Proportion of objectives/benchmarks measured within general content area</td>
</tr>
<tr>
<td><strong>SEC (Porter et al., 2001)</strong></td>
<td><em>Content Match</em></td>
<td>Match items to content &amp; cognitive areas</td>
</tr>
<tr>
<td></td>
<td><strong>Expectations for Student Performance</strong></td>
<td>Compares cognitive demands of curriculum and assessment</td>
</tr>
<tr>
<td></td>
<td><strong>Instructional Content</strong></td>
<td>Compares what is taught with what is tested.</td>
</tr>
</tbody>
</table>
I see alignment studies as one aspect of validity evidence based on test content.

Evidence regarding the degree to which the content of the test is congruent with the testing purpose.

But, alignment related to instruction may also provide validity evidence based on testing consequences.
So, what are the important questions to ask to evaluate the appropriateness of statewide (NCLB) or other educational tests (e.g., district or school formative assessments) for achieving their intended purposes?
Content Validity Questions

1. Do the *test specifications* represent the knowledge and skills specified in the appropriate *curriculum frameworks*?

2. Does the test content sufficiently represent the *test specifications*?

3. Does the content sufficiently represent the *curriculum framework*?

4. Are all items *relevant* to the curricular domain?

5. Are any items potentially biased against certain types of students?

(c) Stephen G. Sireci, 2010
Content Validity Questions

6. Is the content sufficient for providing the information desired, given the testing purposes?
Aligning Curriculum, Instruction, & Assessment
Additional Alignment Questions

7. Has the mandated curriculum had an effect on instruction?
   – Would need to be evaluated over time

8. Are teachers better trained or resourced due to mandated testing?

Provides validity evidence based on testing consequences
In conclusion (1)

● Alignment research can provide important information regarding
  – The degree to which tests are fulfilling their purposes
  – The degree to which students’ performance can be interpreted with respect to the curriculum
  – How an assessment should be changed to better meet its goals
  – Students’ opportunity to learn
Conclusions (2)

- There are many ways to evaluate alignment.
- To select the best method, or piece of a method, identify your goals
  - State
  - District
  - Classroom
- Alignment should be an important characteristic of a statewide testing process.
Alignment Studies

- Provide data that can be combined with the priorities of educational stakeholders to guide changes in assessments, standards, and/or instruction.
- Help interpret students’ test scores
- Can provide important professional development

(c) Stephen G. Sireci, 2010
“Better aligned goals and measures of attainment of these goals will increase the likelihood that multiple components of any district or state education system are working toward the same ends” (Webb, 1997, p. 2).
Suggestions planning/conducting study

1. What level of alignment?
   - State (NCLB), District, School, Class

2. Define alignment needs/goals
   - NCLB accountability?
   - Instruction/Assessment
   - Interpreting students’ performance

3. Select method, or pieces of methods that will provide needed data.

4. Provide information back to interested parties (documentation)

5. Study should be *independent*, if needed

(c) Stephen G. Sireci, 2010
Closing remarks

- Much work to do, but good news is there are methods and research available to help us.
- Thanks to Gil Andrada and CT DOE for invitation, and you for your attention.

sireci@acad.umass.edu