Assessment Research in an NCLB Climate

CERA Annual Meeting
November 2002
Richard S. Brown
University of California, Irvine
No Child Left Behind Act

- Variously referred to as:
  - The No Psychometrician Left Behind Act
  - The Psychometrician Full Employment Act

- Puts assessment at the heart of school accountability

- Generates tremendous opportunity for assessment research
Some of the Assessment Implications of NCLB

- Standards-based accountability
- States and districts must issue report cards to the public (02-03)
- State participation in NAEP (02-03)
- Mandatory reading and math testing of all students in grades 3-8 (05-06)
- Must demonstrate “adequate yearly progress” for all student groups
Standards Based Accountability Systems

- Curriculum & Instruction
- Professional Development
- Assessments
Assessment Areas of Opportunity

- Curriculum & Instruction
  - Alignment
  - Research-based practices
- Professional Development
  - Assessment literacy
- Assessments
  - Instructional sensitivity
  - Measuring adequate yearly progress
Alignment

- Porter (2002)
  - “Alignment is the core idea in systemic, standards-based reform”
  - Developing “alignment indices”
  - “There is no quantitative index of degree of alignment or any ability to compare degree of alignment across states or among assessment standards and textbooks”

- Thus, opportunities to develop such indices are present
Alignment

District

Achievement

Instruction

Assessment

Standards

State

Porter (2002)
Alignment

- Not limited to test/standards alignment
- Alignment also must involve an instruction component
- Alignment studies could also address other uses of assessment data
  - Standards for Success (S4S) studies – articulation across/among systems
To investigate the alignment between state mandated exams and key knowledge and skills (KKS) needed for University success.

Much alignment research to date has focused on the alignment between state tests and state standards – not on the key knowledge and skills needed for University success.
Key Knowledge and Skills

- Developed via S4S National Conversations (Conley, 2001).
- KKS for Mathematics included 89 objectives in 6 standards
  - Computation, Algebra, Geometry, Trigonometry, Math Reasoning, Statistics
- KKS for English/Language Arts included 66 objectives in 4 standards
  - Writing, Reading & Comprehension, Research Skills, Critical Thinking
Methodology

- Based on Webb (1997, 1999)
- Four criteria for content focus alignment
  - Categorical Concurrence
    - The extent to which the test items address the objectives of the key knowledge and skills
  - Depth of Knowledge Consistency
    - The agreement in terms of cognitive complexity between test items and the key knowledge and skills
Methodology (cont.)

- Four criteria for content focus alignment (cont.)
  - Range of Knowledge Consistency
    - The percent of objectives within each standard addressed by the assessment items
  - Balance of Representation
    - The distribution of assessment items across addressed objectives within each standard
Rating Tasks

- Raters rate Standards and Assessment Items in terms of Depth of Knowledge (Cognitive Demand)
  - Depth of Knowledge Rating Categories
    - Recall and Recognition
    - Skills and Concepts
    - Strategic Thinking
    - Extended Thinking

- Raters rate Categorical Concurrence between Standards and Assessment Items
Raters

- Twelve raters participated
  - 6 in Mathematics
  - 6 in English/Language Arts
- Recruited from secondary and post-secondary institutions
Results

- Reliability of Raters
  - Generally good reliability among raters
    - For KKS depth of knowledge ratings (reliability coefficients > .80 with 6 raters)
    - For both Math tests, reliability coefficients > .80
    - For ELA tests, reliability coefficients lower (.70-.80)
Alignment Results

- Depth of Knowledge Consistency Summary
  - Good Depth of Knowledge Consistency for all standards for both state assessments in Mathematics
  - Good Depth of Knowledge Consistency for all standards for both assessments in English/Language Arts
Alignment Results

Categorical Concurrence Summary

- Good categorical concurrence for both state assessment in Mathematics for most standards
- Poor categorical concurrence for both state assessments in ELA for most standards
  - Only Reading & Comprehension standard had sufficient number of items for each state assessment
Alignment Results

- Range of Knowledge Summary
  - Weak range of knowledge for most standards for both state assessments in Math
  - Weak range of knowledge for most standards for both state assessments in ELA
    - Only Writing standard had a majority of the objectives addressed on one exam
    - Weakness primarily due to limited number of items on these exams
Alignment Results

- Balance of Representation Summary
  - Good balance of representation for all standards for both state assessments in Math
  - Good balance of representation for most standards for both state assessments in English/Language Arts
    - Only Reading and Comprehension for one exam did not meet the balance index criterion
## Summary of Results

<table>
<thead>
<tr>
<th>Alignment Criteria</th>
<th>State A Math</th>
<th>State A ELA</th>
<th>State B Math</th>
<th>State B ELA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical Concurrence</td>
<td>Good</td>
<td>Poor</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>Depth of Knowledge</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Range of Knowledge</td>
<td>Weak</td>
<td>Weak</td>
<td>Weak</td>
<td>Poor</td>
</tr>
<tr>
<td>Balance of Representation</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>
Conclusions

- The two state assessments aligned comparably to the KKS
- Range of Knowledge and Categorical Concurrence alignment is difficult to obtain with few assessment items and many standards/objectives
- Consolidation of standards/objectives and/or more assessment items could enhance the coverage of the standards/objectives
Conclusions

- More research in the area of alignment across systems is needed
  - S4S continuing with analysis of about 30 state assessments
  - Similar analyses with other post-secondary institutions (e.g., State school systems, Community Colleges) and other assessments (e.g., placement exams, college entrance exams) should be conducted
  - Also alignment across grade spans (e.g., middle to high school) could be investigated
Research Based Practices

- Educational decisions should be made based on “scientifically based research”
- Limited number of rigorous studies currently exist
- Opportunities to provide local evidence of effectiveness
Research Based Practices

- Slavin (2002)
  - “With a robust research and development enterprise and government policies demanding solid evidence of effectiveness behind programs and practices in our schools, we could see genuine, generational progress instead of the usual pendulum swings of opinion and fashion.”

- Tremendous opportunity for education researchers to have a impact
Over a decade ago, it was suggested that assessment knowledge was important - and becoming more so - and that many educators did not possess sufficient assessment literacy (Popham & Hambleton, 1990)
Assessment Literacy

- More recently, Trevisan (1999)
  - Survey of assessment/measurement requirements for administrators in all 50 states
- Key Findings:
  - Most states still do not require demonstrated assessment knowledge or skills for certification
Assessment Literacy

- Stiggins (1995) described barriers to assessment literacy
  - Fear of assessment and evaluation
  - Insufficient time to assess well
  - Public perceptions of the state of assessment practices
- How to remove barriers?
  - Professional development
Assessment Literacy

- Popham’s admonition to assessment specialists (2002)
  - Due to past sins of omission, a duty (penance) to speak out and educate test users and policymakers about assessment issues is upon us.
Instructional Sensitivity

- Tests’ responsiveness to learning gains
  - Linn & Baker (2002) demonstrate concern for state assessments’ ability to demonstrate learning gains when present

- Issues at High School level
  - Motivation
    - Local district example of DST vs. CST performance
  - Over-testing
Measuring AYP

- AYP Requirements of NCLB
  - All students to improve performance to achieve “proficient” status within 12 years
  - Sub-groups must also demonstrate AYP
- “Safe Harbor” provision
  - If a school makes a 10% reduction in the proportion of one of their student sub-groups rated as not proficient and that group also makes progress on one or more academic indicators, the school will be considered to have met AYP for that year.
Measuring AYP

- Current issues in CA
  - Existing API targets
    - Less stringent than NCLB requirements
    - 5% of distance to goal
    - 80% of school-wide target for sub-groups
Measuring AYP

- Sustainability of improvements
  - Volatility of test scores
    - Kane & Staiger (2001)
    - Linn & Haug (2002)
    - Arce-Ferrer, Frisbie, & Kolen (2002)
  - CA past three years (1999-2001)
    - Rogosa/CDE reports
      - Stairstep vs. See-saw
    - PACE report
Sustained Growth Outcome Summary by School Type for Initially Low Performing Schools

- Elementary Schools
- Middle Schools
- High Schools

Percent of Schools Meeting Criteria:

- API Gain: 88.40%, 71.40%, 56.30%, 43%
- School Target: 34.10%, 20.30%, 3.60%, 2.90%
- Sub-Groups: 8.40%, 3.60%, 2.90%
- Awards: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%
Sustained Improvement

- The extent of sustained improvement realized depends on how the issue is conceptualized.
- However conceptualized, less at progressively higher levels along the system.
- Predictors of initial achievement are not good predictors of sustained improvement.
### Sustained Improvement and Demographic Variables

<table>
<thead>
<tr>
<th>Low performing Elementary Schools</th>
<th>API Change in API Score</th>
<th>School Growth Targets</th>
<th>CI Targets</th>
<th>Award Eligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Free/Reduced Lunch</td>
<td>-.90** (n=4847)</td>
<td>.020 (n=2289)</td>
<td>-.040 (n=2410)</td>
<td>.050* (n=2410)</td>
</tr>
<tr>
<td>Percent ELL</td>
<td>-.77** (n=4847)</td>
<td>.085** (n=2280)</td>
<td>-.016 (n=2399)</td>
<td>.142** (n=2399)</td>
</tr>
<tr>
<td>Percent Mobility</td>
<td>-.12* (n=4847)</td>
<td>.057** (n=2291)</td>
<td>.018 (n=2411)</td>
<td>.012 (n=2411)</td>
</tr>
<tr>
<td>Percent Fully Credentialed Teachers</td>
<td>.57** (n=4847)</td>
<td>-.084** (n=2291)</td>
<td>-.031 (n=2412)</td>
<td>-.093** (n=2412)</td>
</tr>
</tbody>
</table>
Assessment Research Opportunities

- Explore research opportunities at local levels
  - Conduct alignment studies across all levels and activities, including instruction
  - Conduct instructional sensitivity studies, particularly for locally developed measures
  - Conduct rigorous evaluation studies
  - Involve district/university collaboration in these efforts
Assessment Research Opportunities

- Enhance professional development opportunities
  - Assessment literacy workshops for district administrators and teachers
  - Information sessions with school board members
  - Dissemination of research finding to practitioner audience
Assessment Research Opportunities

- Explore locally developed measures
  - Develop benchmark testing, district developed measures, etc.
  - Provide alternative/additional opportunities to demonstrate school improvements
  - Enhance/explore assessments’ instructional sensitivity
Assessment Research Opportunities

- Develop and improve longitudinal student databases and information systems
  - Analyze these data to help guide instructional decisions and target interventions
  - Work with districts to help establish these systems and convey the need for them
Assessment Research Opportunities

- Identify systemic influences on sustained improvement
  - Model influences on demonstrated sustained/consistent improvement where identified
  - Investigate school/classroom practices of schools showing sustained improvement