Challenges to Developing and Implementing Instructionally Supportive Local Assessment Systems_2

Brian Gong

Center for Assessment

Kim Ferguson

Cutting Edge Consulting

June 24, 2003

CCSSO Conference on Large-Scale Assessment

www.nciea.org

Using Local Assessment Systems for Accountability:
Issues of Comparability and Flexibility

General background:

- Some states would like to explore how local assessment information might be used in school accountability systems (e.g., NCLB)
- We are learning about the issues involved, and working on developing possible solutions
  - Interpretation to standards ("what is proficient")
  - Representation and aggregation
  - Design for adequate "comparability"
  - Feasibility
  - Developing local capacity

Comparability and Flexibility

- What comparability do we want, and why?
- How much comparability/flexibility is needed to make good assessment and accountability decisions?
- What are some ways local assessment and accountability systems can be designed to provide appropriate comparability and flexibility?

Assessment and Accountability

- Assessment: how to get information about performance

- Accountability: what happens as a result of that performance

State assessment and accountability systems

- Exist with other systems of assessment
  - Local assessment
  - State assessments
  - Other: NAEP, NRTs, SAT/ACT, etc.
- Exist with other systems of accountability
  - Curricular targets
  - Student grades, report cards, credits, promotion, graduation
  - State ratings, labels, reports, consequences
  - Federally mandated AYP labels, consequences, reports
- Trying to create coherent, powerful comprehensive system by aligning purposes and coordinating elements

Assumptions of Inference in Assessment

- Generalization
  - To performance on similar tasks
  - To construct and performance standard ("what is proficient in math")
  - Across persons and time ("this school is like…", "this person/school will likely be in the future…")
- Because any assessment is a sample, field is developing ways to specify the targets and what and how to sample → GLEs, performance descriptions, Balance of Representation, item specs, test specs, reliability/generalizability, aggregation
- Comparability over occasions, persons, time → Standardization, empirical monitoring/moderation, judgment

Assumptions of Accountability

- Trying to identify the "right" schools for labels, ratings, consequences (e.g., assistance, sanctions, rewards)
- Compare schools to criteria → need valid model and criteria that are appropriate for every school
  - Said was "In Need of Improvement" but wasn't (Type I error)
  - Said "Was OK" but truly was INOI (Type II error)
- Reliability – changes due to more than error (measurement, sampling, equating, etc.) and are applicable to future

Possible "Local" design characteristics of LAS (Local assessment/accountability system)

- Local definition of targets: content & performance standards

- Local definition of sampling plan, e.g., Balance of Representation or accountability weighting

- Local determination of aggregation (combination with other tasks or other information)

"Local" design characteristics of LAS - continued

- Local development of assessments
- Local choice of assessments (common at district, school, teacher, student levels; change over time)
- Local choice of administration timing
- Local administration oversight
- Local scoring
- Local reporting and interpretation
- Local determination of uses and local action

"Local" design characteristics of LAS – continued

- Local context for curriculum and instruction (including "teaching to the test" and other test preparation)

- Local decisions about motivation

- Local decisions about resource allocation (e.g., professional development for LAS, distribution of excellence in teachers/administrators)

Means for comparability

- Same everything for everyone
- "Parallel" and "equated" for everything
- "Same scale" -> same construct, equal weights
- Same standard – same definitions
- Same standard – planned development process or post hoc analysis (e.g., CCSSO alignment study)
- Same standard – review of implementation process (e.g., Queensland monitoring)
- Same standard – reference to external anchor (e.g., state's CRT, e.g., MA)
- Same outcome – regression or other empirical adjustment

How "common" do tasks need to be to incorporate local assessment into school accountability system?

Depends on what level of assertion and what type of evidence is acceptable

- Define targets – content standards (and performance standards, and administration guidelines)
  - Example: Writing portfolios in KY, VT
- Define sampling plan (representation) and aggregation rules (and pool of assessments)
  - Example: Wyoming "Body of Evidence"
- Define process and criteria for sampling plan, assessment development and administration (and monitoring criteria and process)
  - Example: Maine Local Assessment System (MA stu. accy)

What about replacement of tasks in LAS?

- For student accountability, multiple opportunities to retest are standard practice – addresses reliability and possible Type I error (Type II error rates rarely discussed)

- For school accountability, multiple student retests are not common (addressed through larger number of students); multiple years of data may be used (to address sampling error)

- In LAS, often want to combine purposes of student and school accountability. Need to consider:
  - Inference/decision tied to what (representation);
  - Will replacement change decision when combined with other evidence (representation and aggregation);
  - When/under what conditions can evidence be replaced (e.g., conclusion about student performance at any-point-in-time vs. end-point of grade or school vs. "best work")

Challenge

- Require local assessment and accountability systems to carry more weight in our accountability judgments and educational uses

- Develop new sensibilities about what makes things "comparable" and what is enough comparability

- Develop understanding of how assessment and accountability systems are put together

- Try new models, implement in real contexts

Implementation Strategies to Ensure Success of Local Assessment Systems

Building Local Capacity

- Peer Review

- Consortium of Districts

- Cadre of Experts

Peer Review:

Peer review process to evaluate and critique local assessment systems for the five required criteria:

- ❖ alignment,
- ❖ consistency,
- ❖ fairness,
- ❖ standard-setting and
- ❖ comparability.

Peer Review Process to evaluate local assessment systems:

Criteria, Score, and Rationale

Alignment: Satisfactory

The district plan includes evidence of two-way alignment and adequate sampling of the full range of standards in Mathematics, and Physical Education. Our assumption is that the district plan/procedure to ensure two-way alignment, adequate sampling of the full range of standards, and alignment to cognitive levels represented by the performance standards for other content areas is in place but not included in the plan.

Consistency: Progressing

There is evidence that procedures to assure consistency are planned, but have not been implemented yet.

Peer Review Evaluation Criteria, Score, and Rationale

Fairness: **Satisfactory**

The district presents evidence that it has procedures and a system in place to ensure that assessment items and tasks are not biased against any subgroups. Accommodations, alternate assessments, multiple opportunities, and different assessment formats and strategies are mentioned in the plan as well.

Standard-Setting: **Progressing**

There is evidence that the district has a well-articulated plan and a timeline to use a defensible standard setting approach. With implementation of the plan including stakeholders in the standard setting process, this section should move to satisfactory.

Comparability: **Satisfactory**

The district has provided evidence that it has procedures to ensure comparability such as a systematic plan for revising and replacing items, district wide training, common rubrics, anchor papers and exemplars.
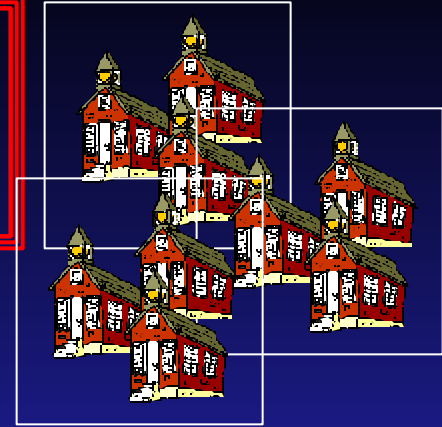
Value Added of Peer Review:

❖ deeper understanding of assessment systems, approaches, and solutions.

❖ greater "assessment literacy" in <u>district personnel</u> including teachers, building administrators, and curriculum directors.

❖ greater consistency and coherence of assessment systems in local districts.

Consortium of Districts:

- develop and field test common assessments in mathematics, science, language arts, and social studies

- collect student work to determine anchors over a three year process.

- build assessment literacy.

Cadre of Content Experts:

❖ Build scoring packets and anchor papers for district use;

❖ Deliver regional training sessions to build consistency in local scoring;

❖ To implement strategies for embedding assessments into local curriculum;

❖ To develop unit standards-based units of study.

Value added of Cadre of Experts:

- Building long term capacity in the system

- Building assessment literacy;

- Strengthening link between assessment, curriculum, and instruction;

- Statewide collaboration between content experts

Building capacity by empowering professionals with the expertise they need to successfully implement local assessment systems through:

- Peer Review

- Consortium of Districts

- Cadre of Experts

All together now!

For more information

The Center for Assessment

www.nciea.org

bgong@nciea.org

Kferguson@cutting-edge-consulting.com

Check website for information on 2003 RILS conference sponsored by the Center and WestEd

- ◆ What states are doing with *No Child Left Behind*
- ◆ Reliability and *NCLB*
- ◆ Alignment and *NCLB*