

Documenting Validity of Accountability Systems: A progress report

Brian Gong
Center for Assessment

Presentation at the CCSSO National Conference on
Large-Scale Assessment
June 20, 2005 San Antonio, TX



Accountability system validity – a new endeavor

- While validity of assessment systems has a long tradition, many states are just embarking on establishing the validity of their accountability systems. From design of indicator systems to accuracy of data elements to improvements in student learning and avoidance of unintended consequences such as narrowing of the curriculum and cheating, states are beginning to document their accountability systems.
- This session will provide an overview of how accountability validity differs from assessment validity, and where you might start practically.



Validity issues are different for assessment versus accountability

	Validity	Reliability
Assessment	Measure the student performance on the construct (e.g., mathematics)	Consistent scores across occasions, forms (<i>measurement error, equating error</i>)
Accountability	Identify (not) schools that should be identified, consistent with construct and rules; Have desired effects (not undesired)	Consistent accountability decisions (<i>sampling error; design decisions</i>)



Accountability system validity

System design	Execution	Reporting and interpretation	Dept. response	Impact & sustainability
Goals, construct indicators, combination rules, who's included, etc.	Assessment construction & administration; Data in; data processing; data checks	How are results portrayed, communicated, and used	What does Dept. do to improve situation?	Student performance, LEA capacity & actions; community support



Examples of *system design* affecting accountability validity

- What defines a “good” or “bad” school?
- Reliability
- Inclusion



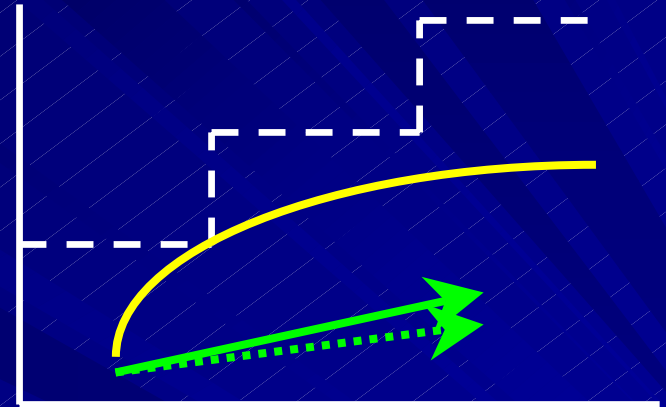
What defines a “good”/“bad” school?

- Status, improvement, growth, other
- How much is “bad enough”?
 - All students, average performance, subgroups
 - NYYYYYYYY versus NNNNNNNNY
 - Taking care of Type I and Type II error
 - Use two-stage systems
- What happens to a “bad” school?
 - Same consequences, no matter the pattern
 - Focused consequences, e.g., sanctions follow subgroup; subgroup x content area



What defines “good”/“bad” school”? - 2

- How much is “good enough”?
 - Status vs. Safe Harbor (and now student growth models)



- Value Tables (see Hill, 2005)

		Year 2		
		BB	B	P
Year 1	P	0	0	100
	B	0	0	100
	BB	0	0	100



Reliability design (vs. validity?)

- Less reliable vs. more reliable
 - Proficient/not vs. ach. levels or scale score
 - Percent proficient vs. index or ss average
 - Smaller subgroups vs. larger size groups
 - Conjunctive vs. compensatory
 - More correlated vs. less correlated
 - More conjunctive decisions vs. fewer
 - Change scores vs. status

See Hill et al. (2000, 2001, 2002, 2003, 2004, 2005)



(In)Exclusion through minimum-n

Percent of passing schools that did not meet minimum-n for SWD subgroup (percent of SWD in state excluded)

State	Minimum-n Size			
	10	20	30	60
1	34 (10)	75 (39)	83 (50)	86 (97)
2	65 (19)	92 (54)	97 (76)	100 (99)
3	53 (11)	82 (41)	96 (74)	100 (99)
4	71 (9)	83 (21)	91 (32)	100 (72)
5	42 (2)	69 (7)	89 (20)	99 (68)

Source: Center for Assessment, Simpson, Marion & Gong, forthcoming



(In)Exclusion through FAY

Mobility – Did not meet “Full academic year” requirement

GRADE	4	8	HS
STATE	2%	3%	2%
LEA	4%	5%	3%
SCHOOL	8%	9%	5%

MOVERS, Percent Free/Reduced-Price Lunch

GRADE	4	8	HS
STATE	87%	84%	80%
LEA	84%	78%	72%
SCHOOL	87%	77%	65%

From Simpson & Gong, forthcoming



Improving validity: Where to start

System design				
Focus on clear purpose Understand options Weigh validity heavily when making compromises Keep improving system design				
Involve your TAC, talk with others (e.g., ASR SCASS), see Accountability Technical Manual Table of Contents (Gong, 2004)				



Improving validity: Where to start - 2

System design	Execution			
	Monitor & progressively fix your weakest areas Make a sustainable system Inculcate shared responsibility with LEAs, others			



Improving validity: Where to start - 3

System design	Execution	Reporting & interpretation	Department response	Impact & sustainability
			<p>Give proportional resources to helping schools improve (more than measuring)</p> <p>Figure out Department role (esp. when to be direct service provider); be coherent</p> <p>Be serious with adults/students in system (telling does not equal teaching; \$ do not equal capacity to change; their motivation & skills really matter)</p> <p>Easier to keep institutions than to rebuild them</p>	



Session Outline

- a) What have you recently put effort towards in documenting or improving the validity of your state's accountability system?
- b) What do you plan to do next?
- c) What lessons have you learned, and what advice would you give to other states and agencies (e.g., USED) regarding improving the validity of school accountability systems?

- Mitch Chester – Ohio Department of Education
- Tom Spencer – Louisiana Department of Education
- Sue Rigney – U.S. Department of Education
- Brian Gong – Moderate discussion with audience



For more information:

Center for Assessment
www.nciea.org



Brian Gong
bgong@nciea.org

P.S. See website for information on our upcoming RILS conference on high school assessment and accountability, Sept. 29-30, 2005

