Between a Rock and a Hard Place: Adjusting AYP Workbooks to increase reliability *and* validity

Brian Gong

Center for Assessment

AERA Annual Meeting Discussion on

NCLB in Year Three: The state of the accountability plans

Montreal, Quebec, Canada April 14, 2005

States' (and USED's) Dilemma

- Current AYP rules will "over-identify" schools, i.e., be politically unacceptable, strain improvement resources, identify many schools that should not be identified in terms of quality, be very unreliable in initial identification, and be very unreliable in not letting schools out (and still not identify many that should be identified)

- The main tools for solving this are inadequate and invalid
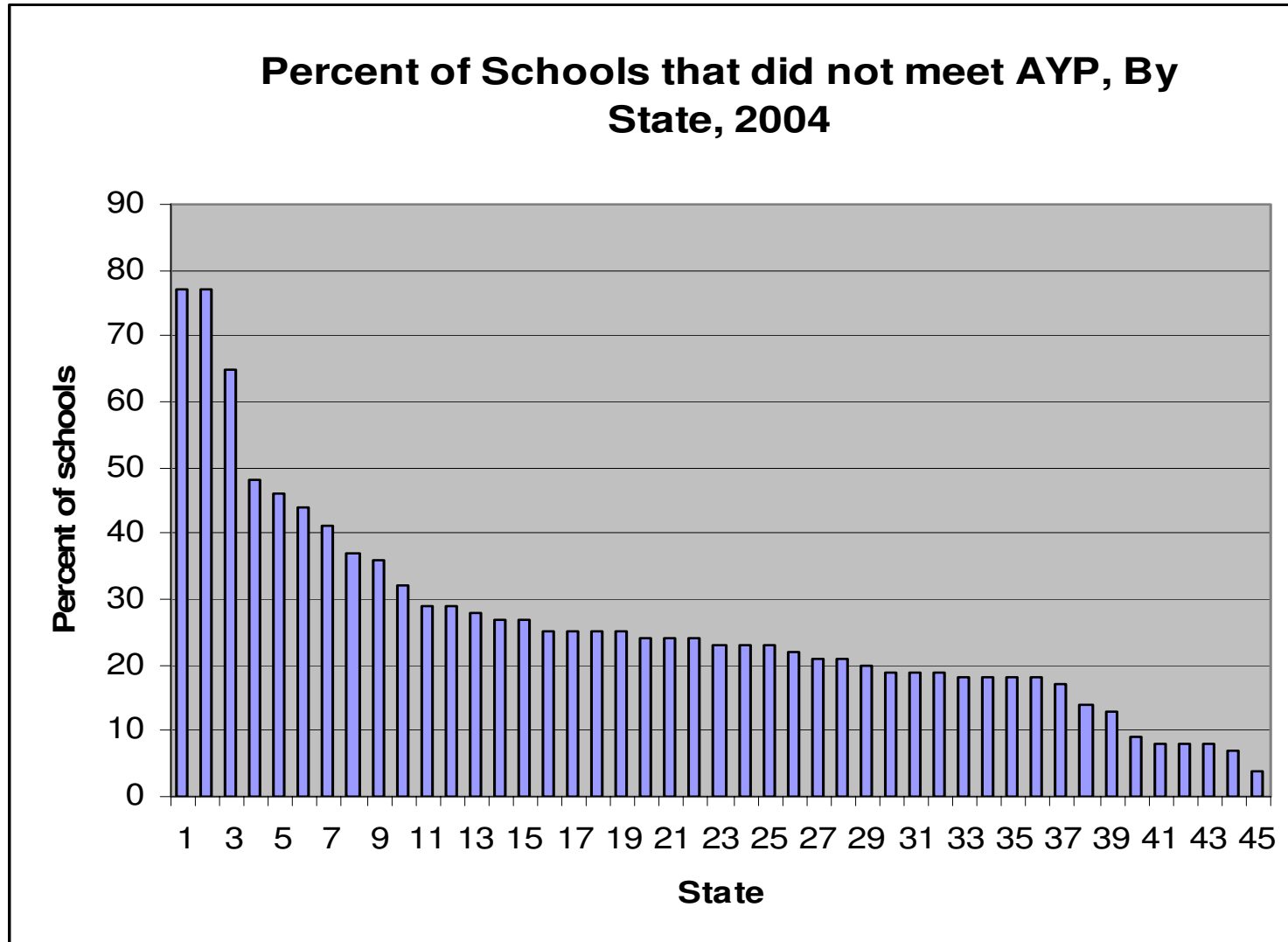
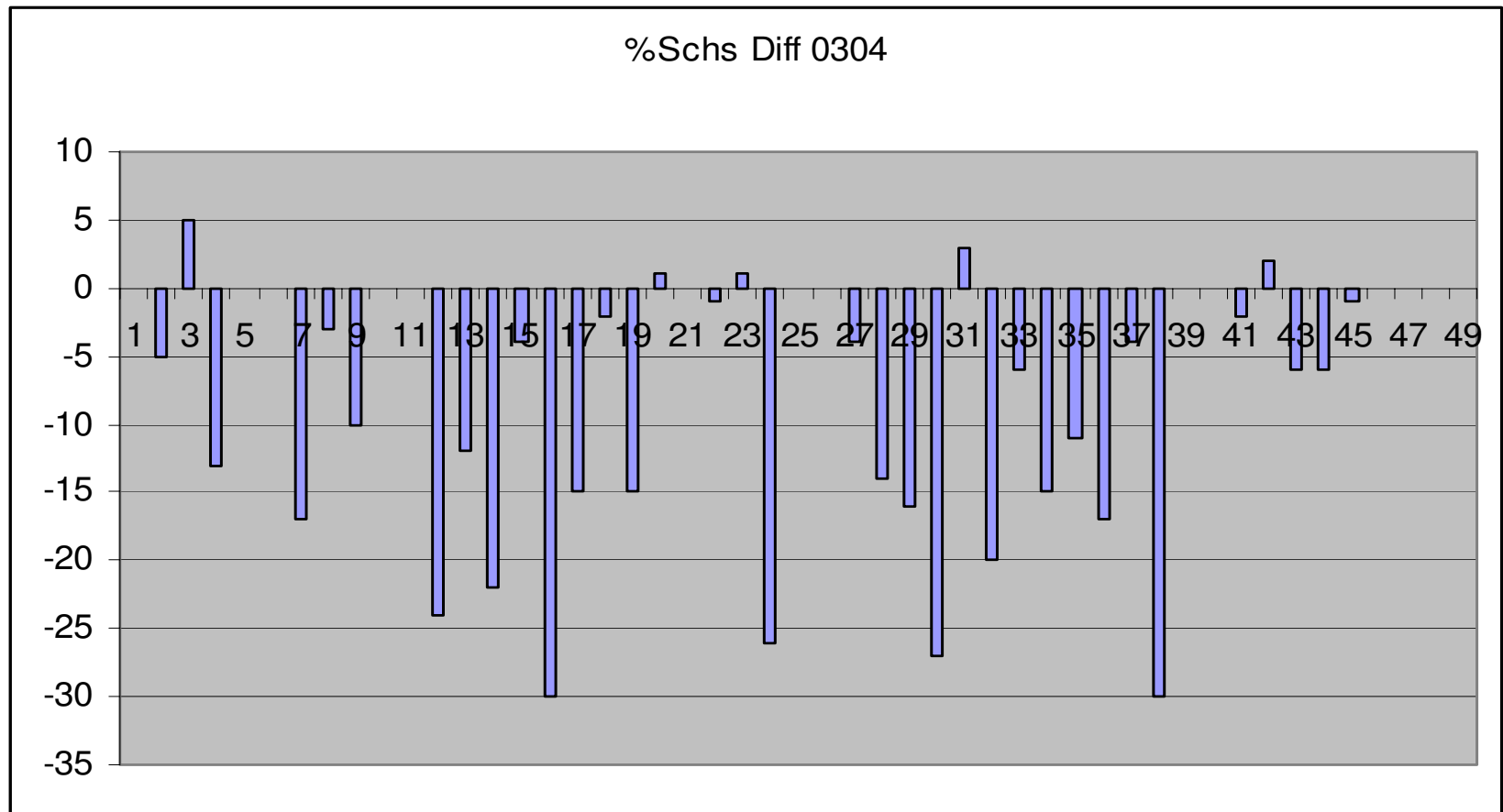# "Over-identification" in the making

Percent of Schools that did not meet AYP, By State, 2004

Real improvements in school performance (no), or…?

5 increase (avg. 1%), 32 decrease (avg. 15%), 9 no comparison, 7 missing from table

%Schs Diff 0304

Four dimensions that affect school identification as not meeting AYP

- Performance
- Inclusion
- Standards
- Reliability

Current Policies That May Reduce Identification of Schools - 1

- Performance improvement
  - Meeting AMOs over time
  - Safe harbor improvement

Current Policies That May Reduce Identification of Schools - 2

- Exclusion
  - Full Academic Year
  - 5% rule on participation
  - Non-participation (medical, etc.)
  - Minimum-n
  - Appeals (!)
- Lower Standards
  - "Proficient for NCLB" = Basic
  - Stair-step increase of AMOs out to 2014 ("balloon payment")
  - 1% Alternate
  - 2% modified ("out of grade")
  - Longer time rates for graduation
  - Identify ELL and SWD in subgroup after classified out of subgroup
  - Graduation rate and other academic indicator bar set low (e.g., 20th percentile school, only 0.1% increase needed)
- Reliability
  - Confidence Interval (on performance and other indicators)
  - "rounding" rules for 95% participation
  - Uniform averaging

Dilemma

- States trying to balance misclassification error
  - Identifying schools as failing to meet AYP that should not be identified (Type I error) and
  - *Not* identifying schools as failing to meet that *should have been identified*

- Most adjustments now decrease Type I error (reduce erroneous identification) but simultaneously increase Type II error (increase erroneous non-identification)

- Many of the most prevalent adjustments (particularly exclusion) undermine validity and do not help reliability much

Example: Exclusion through minimum-n (SWD)

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 & Gong, forthcoming

Solutions to Dilemma – Proposals for "New Flexibility"

- Design to reduce *both* Type I and Type II misclassification
- Focus identification on performance
  - Same subgroup, same content area, two years
  - Define "proficient" as "on track to be proficient" and allow student growth models
  - Allow different but convergent subgroup AMOs

Solutions - 2

- Boost inclusion
  - Use confidence intervals on status *and* safe harbor (99% individual, or even 99% familywise) instead of high minimum-n's

- Most important: Promote two-stage systems that demonstrate reliability and validity (minimize Type I and Type II errors by applying different criteria for Stage I and Stage 2)

Solutions - 3

- Consequence follows subgroup (e.g., if SPED subgroup fails to meet AYP, then SPED subgroup is offered choice and/or supplemental services, not whole school)

- Supplemental services before choice

Other adjustments (longer-term)

- Do research to decide whether SPED should be further differentiated into more than two groups, with growth expectations
- Use combination of research and policy to decide about AMO expectations (major item!)
  - Allow "same slope" models for new cohorts (ELL)
- Support Peer Review of reliability and validity of states' accountability systems
  - validity much more than what was addressed here (see, for example, E. Forte Fast & Hebbler, CCSSO, 2004; Gong, CCSSO, 2004; S. Lane, CCSSO, 2005)
- Fix HOUSE teacher quality regulations, supplemental service providers… (whole system look at NCLB statute)

Summary

- Focus on adjustments that increase the reliability *and* validity of the AYP system
  - Incorporate two-stage systems that can address both Type I and Type II misclassification error
  - Sharpen focus on performance
  - Make consequences more nuanced
  - Do basic research and adjustments to goals
  - Attend to supports and other systemic aspects of quality schooling

For more information:

Center for Assessment

www.nciea.org

Brian Gong

bgong@nciea.org