Keeping the Baby and (Most of) the Bathwater: Mid-course adjustments of NCLB's AYP

Brian Gong

Center for Assessment

NCME Annual Meeting Invited Panel Discussion on

Next: What Should Be Retained, Adjusted, or Scrapped in the Current Federal Education Policy?

Montreal, Quebec, Canada April 12, 2005

Focus: Good, Practical Adjustments

- Address real problems of many states
- *Could be made within regulatory change*; do not require statutory change or scrapping the law
- Are centered on improving validity of accountability decisions
- Don't let the perfect stand in the way of the good

Some AYP credibility problems

1. Too many school identified as not meeting AYP
2. Wrong schools identified/not identified
3. “Safe harbor” not safe
4. Games playing – loss of focus
5. Small offense, big consequence
6. Different offenses, same consequence
7. Wrong consequences
8. Incoherent design, lack of credibility
9. Schools flip in and out of Did Not Meet AYP
10. Unreasonable goals, too fast

Where’s USED?

- USED approving some “fixes” that undermine the intent of the law
- USED silent on really asking for evidence about validity and reliability of states’ accountability systems

BUT…

- Have political window of opportunity with “new flexibility” to make mid-course adjustments

Pressure to “identify the right number” of schools

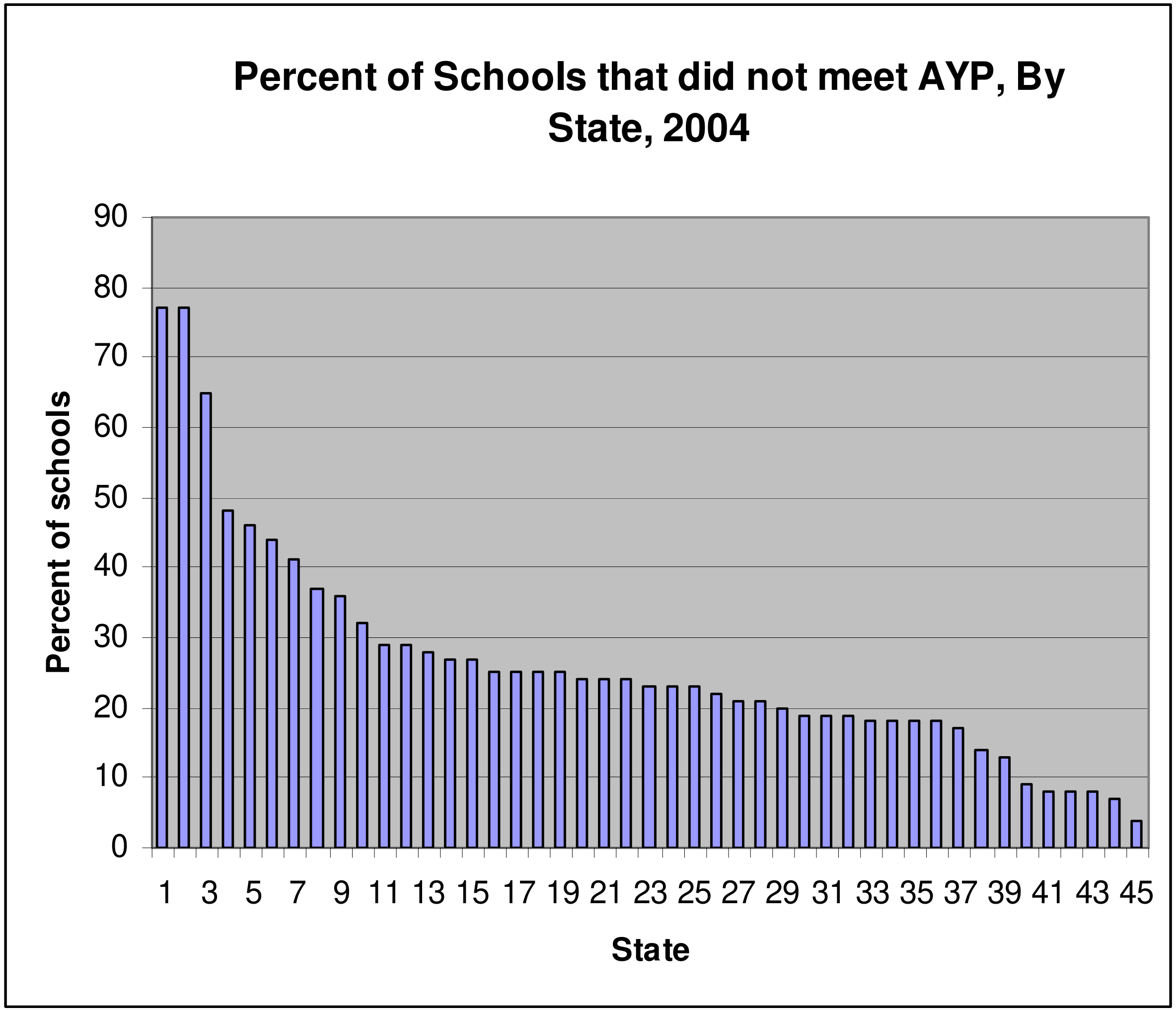


Figure adapted from data published in *Education Week*, “Taking Root,” by Lynn Olson, Dec. 8, 2004, retrieved on 3/7/04 from http://www.edweek.org/ew/articles/2004/12/08/15nclb-1.h24.html

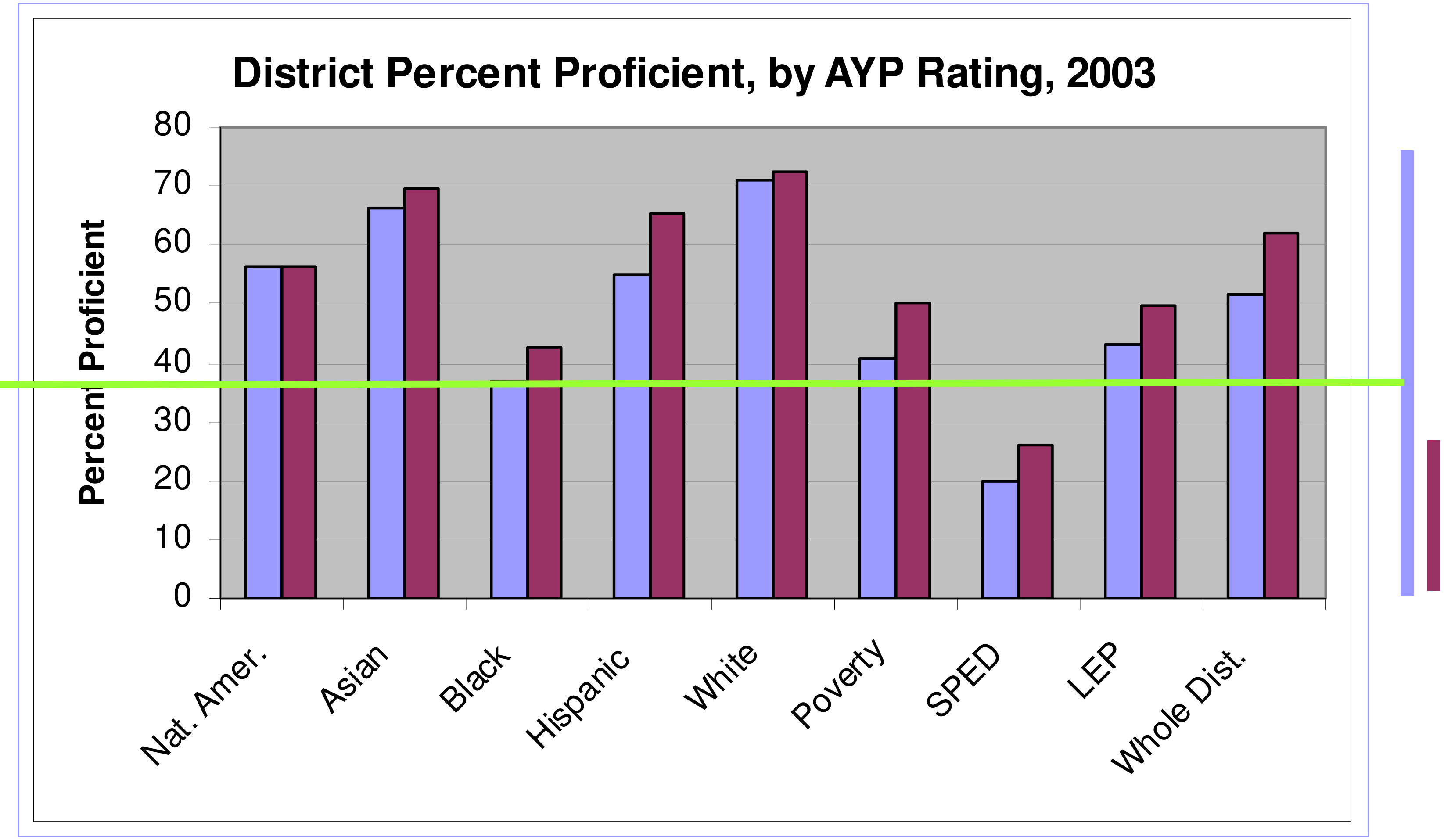
An example: Minimum-n

- Minimum-n size originally intended to help address sampling error and provide some reliability around school decisions, along with the "do not meet two years in a row"
- As threatened by high numbers of schools identified, states and USED have used minimum-n as a way out
  - Approved subgroup minimum size increasing to well beyond 30, plus proposed percentages (e.g., 15% of total student body)

Increasing Minimum-n: "Lose the baby and bathwater" solution

- Statistically inferior to use of confidence intervals
- Biased against large, diverse schools
- Protection against decision inconsistency for status has diminishing returns
- Demonstrably insufficient to guard against unreliability in safe harbor decisions
- Can have tremendous impact on invalidity of AYP design

AYP biased by minimum-n



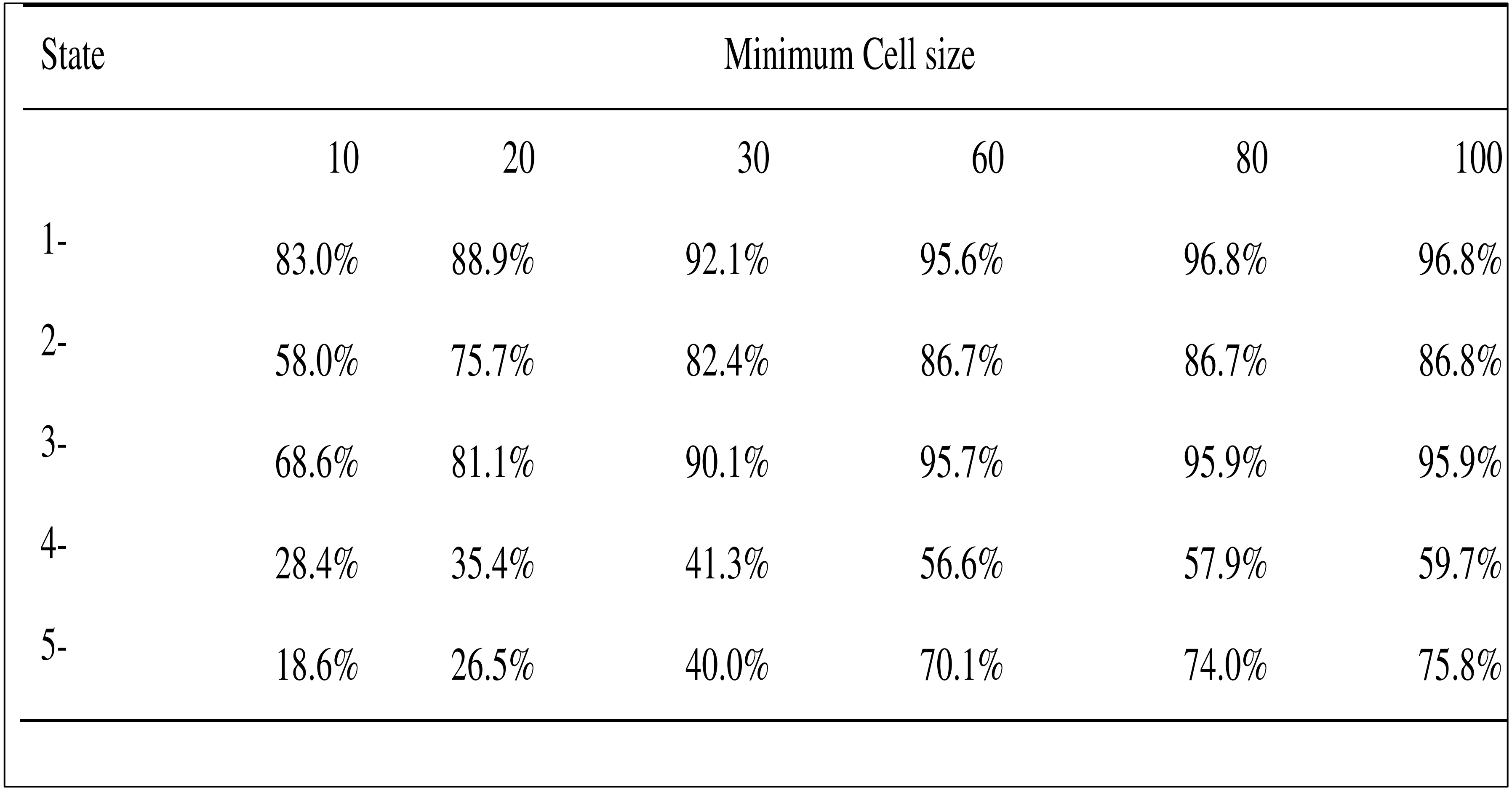
Impact of Increasing Minimum-n – 1

current AMOs & n-sizes, five states, only SPED

State	Passed: School-As-A-Whole (% of schools)	Passed: Special Education (% of schools)	Passed * (% of schools)	Passed but Lacking Minimum *n* in Special Education (% of passing schools)
1- (*n* = 277)	96.8 %	75.3 %	92.2 %	82.7 %
2- (*n* = 1283)	86.8%	34.2 %	79.4 %	94.0 %
3- (*n* = 1112)	95.9 %	49.3 %	87.9 %	90.4 %
4- (*n* = 440)	61.8 %	13.6 %	46.5%	93.5 %
5- (*n* = 723)	78.8 %	10.1 %	50.9 %	92.1 %

Impact of Increasing Minimum-n – 2

Percent of schools meeting AYP

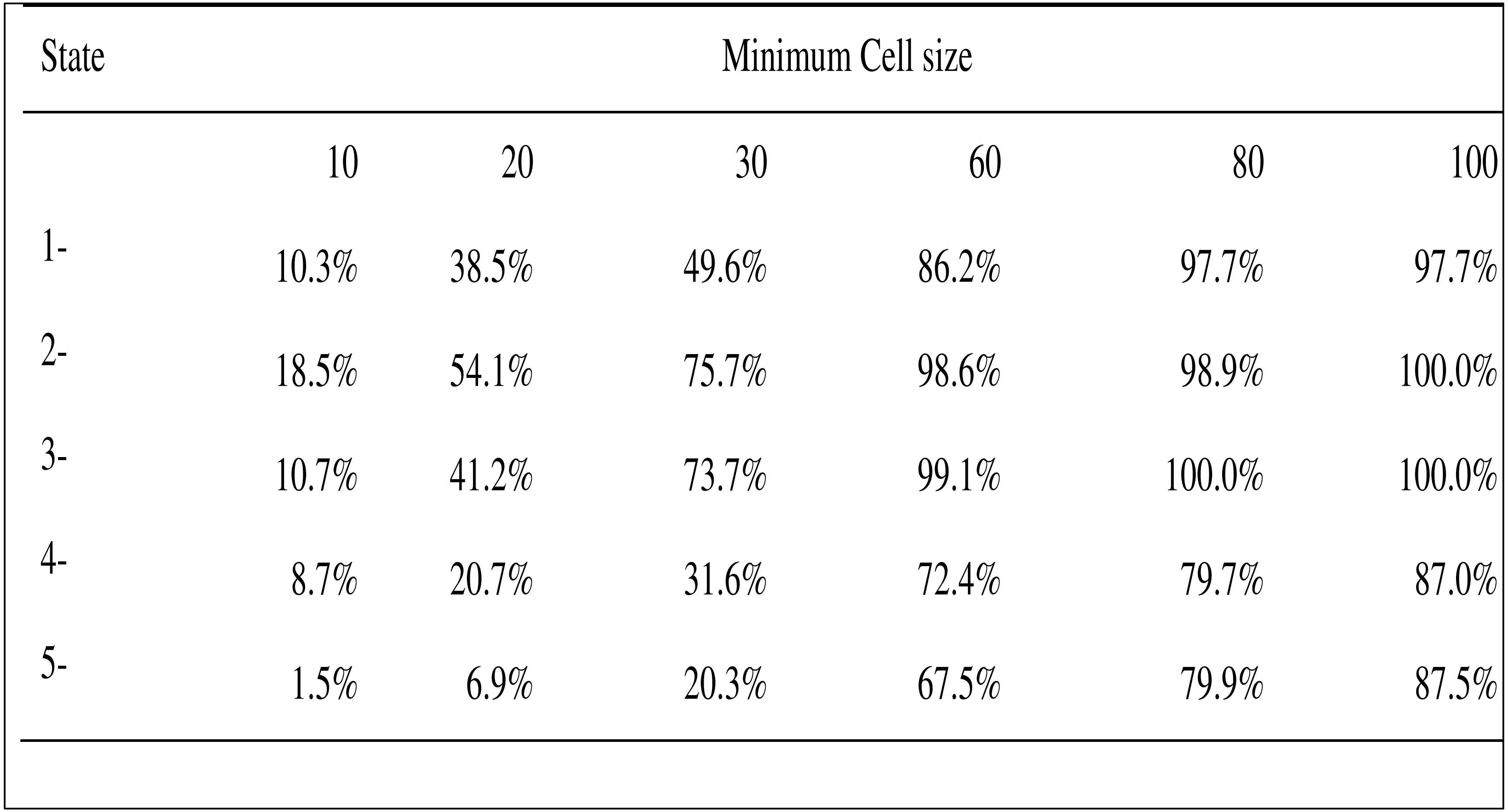


State	Minimum Cell size					
	10	20	30	60	80	100
1-	83.0%	88.9%	92.1%	95.6%	96.8%	96.8%
2-	58.0%	75.7%	82.4%	86.7%	86.7%	86.8%
3-	68.6%	81.1%	90.1%	95.7%	95.9%	95.9%
4-	28.4%	35.4%	41.3%	56.6%	57.9%	59.7%
5-	18.6%	26.5%	40.0%	70.1%	74.0%	75.8%

Impact of Increasing Minimum-n – 3
Percent of passing schools not meeting minimum-n for SPED

State	Minimum Cell size					
	10	20	30	60	80	100
1-	34.3%	75.4%	83.1%	97.1%	99.6%	99.6%
2-	65.0%	91.9%	97.3%	100.0%	100.0%	100.0%
3-	53.1%	81.9%	95.8%	100.0%	100.0%	100.0%
4-	70.6%	83.4%	91.3%	99.7%	100.0%	100.0%
5-	42.4%	69.0%	88.7%	99.3%	99.8%	99.9%

Impact of Increasing Minimum-n – 4
Percent of SPED students in the state excluded



State	Minimum Cell size					
	10	20	30	60	80	100
1-	10.3%	38.5%	49.6%	86.2%	97.7%	97.7%
2-	18.5%	54.1%	75.7%	98.6%	98.9%	100.0%
3-	10.7%	41.2%	73.7%	99.1%	100.0%	100.0%
4-	8.7%	20.7%	31.6%	72.4%	79.7%	87.0%
5-	1.5%	6.9%	20.3%	67.5%	79.9%	87.5%

Impact of Increasing Confidence Intervals

Percent of schools identified as meeting AYP (status)

State	Confidence Interval Size				
	NONE	75	90	95	99
1-	89.8%	90.9%	92.7%	93.0%	94.5%
2-	70.6%	76.5%	80.6%	83.0%	86.2%
3-	83.1%	86.0%	88.5%	90.0%	91.8%
4-	37.7%	43.0%	47.2%	49.6%	55.2%
5-	45.8%	48.3%	51.4%	52.6%	56.4%

Adjustment 1: Approve high confidence intervals on status and safe harbor

- Do not approve high minimum-n sizes for subgroups, if allowed high CIs (99%) on *both* status and safe harbor
  - 95% on each test avg. equivalent to 90% on family of decisions across multiple conjunctive decisions (see Hill & DePascale, 2003)

Make safe harbor more valid

- Look for school improvement reliably over one year, two years, or three years
  - With confidence interval, may not be able to decide reliably with one year of data, but could with two or three
    - School had 10% proficient in Year 1; safe harbor target was 19%. With 99% CI couldn’t identify school as not meeting AYP
    - In Year 2, safe harbor target is 27.1%; in Year 3, safe harbor target is 34.4%

Make minimum-n more valid

- If not using a confidence interval, then minimum-n creates a sharp break
  - School with 30 students is in, school with 29 students is out, no matter their performance, e.g., school with 5 students of 29 proficient declared “Meets AYP” by virtue of minimum-n
  - Using an optimizing calculation—or “benefit of the doubt” approach—regarding minimum-n, could make reliable judgments about these schools
    - School in example could have a maximum of 6 students proficient – would it meet the AMO (with a CI)?

Other adjustments

- Same content area, same subgroup to be identified as not meeting AYP (like districts have to miss in same content area by grade span "subgroups")
- 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)
- Promote two-stage systems (design for reliability and validity, minimize Type I and Type II errors)

Other adjustments (longer-term)

- Do research to decide whether SPED should be further differentiated into more than two groups, with growth expectations
- Allow student longitudinal growth models for school accountability that meet the principles of an ultimate goal of proficiency for all students within a reasonable timeframe (Allow “on track to be proficient” to meet AYP; support index systems for movement towards proficiency)
- Consider fixes to conjunctive unreliability

Other adjustments – 2

- Decide about AMO expectations
- 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… (whole system look at NCLB statute)

Summary

- Focus on adjustments that increase the validity of the AYP system
  - Solve real problems that don’t make sense to schools and public (like “small offense, large consequence” and “different offense, same consequence” as well as political problems (like “over 80% of districts identified as not meeting AYP”)

For more information:

Center for Assessment
www.nciea.org

Brian Gong
bgong@nciea.org