PROMOTING EQUITY IN STATE ACCOUNTABILITY SYSTEMS: A DEEPER LOOK INTO THE WAIVERS

CCSSO National Conference on Student Assessment, June 20, 2013

Wes Bruce, Indiana Department of Education
Chris Domaleski, Center for Assessment
Judy Park, Utah State Office of Education
Marianne Perie, University of Kansas
Allison Timberlake, Georgia Department of Education
Agenda

- Introduction, Chris Domaleski
- Summary of state waiver approaches, Marianne Perie
- Consolidated subgroups in Indiana, Wes Bruce
- Georgia’s College and Career Ready Performance Index, Allison Timberlake
- Including growth in Utah’s Comprehensive Accountability System, Judy Park
- Opportunities and Threats, Chris Domaleski and Marianne Perie
- Questions
Promoting Equity

- Educational equity
  - Based on the principles of fairness and justice
  - Refers to the aim of promoting improved academic achievement for low performing students
- The goals of Title I and IDEA have been to reduce academic achievement gaps
  - Are the waivers promoting equity?
  - What can we learn from state innovations?
PROMOTING EQUITY IN STATE EDUCATION ACCOUNTABILITY SYSTEMS

Marianne Perie
Center Educational Testing and Evaluation
Presentation at the National Conference on Student Achievement on June 21, 2013
NCLB Flexibility Waivers

• Many states took immediate advantage of the Flexibility Waivers announced in 2011
  • 11 states applied under the first submission window in November of 2011
  • 26 more states plus DC applied in the second window of February 2012
  • An additional 7 states applied in the third window of September 2012
  • Three more states applied in February 2013.
Where are we now?

- As of June 6, 2013 at 4:37 pm…
  - 37 states plus the District of Columbia have been approved
  - 7 states and Puerto Rico applied in September 2012 or earlier and are still under review. This number includes Iowa that was initially rejected but is back under review.
  - 3 states initially applied but then withdrew their applications
  - California’s proposal was the only one flat out rejected.
State Status

Approved
Rejected
Under review
Did not submit

+AK & HI
Areas of Difference

• Conducted an analysis of approved waivers in August 2012. At that time, 34 states had received approval.

• Focused on issues of equity and how states varied in their approach to:
  • Student groups
  • Achievement gaps
  • Growth
  • Combining indicators
Including Student Groups

• Crux of public comment on equity
• Achievement of all ESEA subgroups must still be calculated and reported BUT accountability decisions can be made without considering each group separately.
• States tended to do one or both of the following:
  • Lower minimum n-size
  • Combine student groups in one of several ways
Consolidated Subgroups

Majority of states focused on a specific percentile. “Other” includes combining only 2-3 groups that are pre-determined or are based on lowest-performing in school.

<table>
<thead>
<tr>
<th>Number of States Using Consolidated Subgroups</th>
<th># states</th>
<th>% states</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>8</td>
<td>24%</td>
</tr>
<tr>
<td>Bottom Quartile</td>
<td>7</td>
<td>21%</td>
</tr>
<tr>
<td>Combine at risk ESEA subgroups</td>
<td>6</td>
<td>18%</td>
</tr>
<tr>
<td>Only combine groups if minimum n is not met</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Bottom 30%</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Below proficiency for growth</td>
<td>2</td>
<td>6%</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>15%</td>
</tr>
</tbody>
</table>
Achievement Gaps

Focal/reference groups

• Lowest subgroup vs. highest subgroup
• Lowest subgroup vs. opposite
• At risk group vs. not at risk group
• Lowest percentile vs. highest percentile
• Lowest percentile vs. all others
• School vs. state

Metric

• Majority of states used percent proficient
• A few used growth gap
• Two used scaled scores
• One used z-scores
Metric to Calculate Achievement Gaps

The vast majority of states focus on the percent proficient for the gap analysis.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Number of States</th>
<th>Percentage of States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent proficient</td>
<td>18</td>
<td>53%</td>
</tr>
<tr>
<td>Growth gap</td>
<td>5</td>
<td>15%</td>
</tr>
<tr>
<td>Scale score</td>
<td>4</td>
<td>12%</td>
</tr>
<tr>
<td>z score</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Index gap</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>None</td>
<td>4</td>
<td>12%</td>
</tr>
</tbody>
</table>
Movement to Growth

Not all states use the growth model in their accountability system, but instead use it in their educator effectiveness calculations.

<table>
<thead>
<tr>
<th>Number of States Using Various Growth Models</th>
<th># states</th>
<th>% states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Growth Percentiles (SGP)</td>
<td>14</td>
<td>41%</td>
</tr>
<tr>
<td>Value added model (VAM)</td>
<td>7</td>
<td>21%</td>
</tr>
<tr>
<td>Categorical model (value table)</td>
<td>4</td>
<td>12%</td>
</tr>
<tr>
<td>Gain score model with vertical scale</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Improvement</td>
<td>2</td>
<td>6%</td>
</tr>
<tr>
<td>Gain score model using Z-scores</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Still deciding</td>
<td>3</td>
<td>9%</td>
</tr>
</tbody>
</table>
Combining Indicators

NCLB

- Conjunctive decision rule
  - Miss one target and you miss the goal
  - Up to 42 targets, depending on size of student groups in school

ESEA Flexibility

- 3 states maintained conjunctive approach
- Many rank ordered schools on each metric with decision rules for categorizing school
- Most created an index
Various Methods for Combining Measures

The vast majority of states created some type of index to combine measures. Many chose to have cut scores to categorize schools as A through F or ★ through ★★★★★. Indices often include status, growth and gap measures and may include other measures such as graduation rate, attendance rate, and school culture variables.

<table>
<thead>
<tr>
<th>Method for Combining Measures</th>
<th>Number of States</th>
<th>Percentage of States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>22</td>
<td>65%</td>
</tr>
<tr>
<td>Rank order each measure separately</td>
<td>5</td>
<td>15%</td>
</tr>
<tr>
<td>NCLB-type conjunctive</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>12%</td>
</tr>
</tbody>
</table>
CONSOLIDATED SUBGROUPS IN INDIANA
Indiana’s Conundrum

2011 AYP Cells

92
515
688
203
146
121
16
0
Indiana AYP Accountability

- Had the same consequences for every school
- Depended on proficiency
- Provided perverse incentives (“bubble kids”)
- 600+ schools not accountable for any group outside of FRL
  - Clearly we were not using the same yardstick for all schools
  - 92 schools with 9 cells  -  16 schools with 33 cells
- Widespread concern about the “fairness” of the state accountability system
  - Capped by AYP (increasing numbers of schools capped)
Theory of Action

• Address the achievement gap
  • No significant gap closing under AYP
• Make accountability the same “size” for all schools
• Focus on a meaningful educational outcome
• Emphasize growth not just status
• If bottom 25% closes gaps other groups gaps will close
Enter the Super Subgroup!

- Bottom Quartile
  - Calculated by grade and content area
  - Comparison group – balance of the school

- Students in that school with the greatest academic needs

- Every school has 25% of their students in this group
  - Except schools with no tested grades and other outliers

- Rolls up to the district level seamlessly

- Uses Student Growth Percentiles as the metric
  - Schools grades can earn bonus in each content area if bottom 25% shows high growth

- Implemented in 2012 accountability for the first time
Results School A

- 431 students 60% FRL, 46% Minority, 71% Prof 2011
  - Mathematics

<table>
<thead>
<tr>
<th>Group</th>
<th>% Prof 2011</th>
<th>% Prof 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>75.0</td>
<td>82.6</td>
</tr>
<tr>
<td>Hispanic</td>
<td>66.8</td>
<td>87.1</td>
</tr>
<tr>
<td>White</td>
<td>80.2</td>
<td>88.5</td>
</tr>
<tr>
<td>FRL</td>
<td>67.8</td>
<td>86.7</td>
</tr>
<tr>
<td>Paid</td>
<td>86.5</td>
<td>88.9</td>
</tr>
<tr>
<td>Bottom 25%</td>
<td>27.8</td>
<td>88.9</td>
</tr>
<tr>
<td>Top 75%</td>
<td>82.5</td>
<td>87.5</td>
</tr>
</tbody>
</table>
Results School H

- 495 students 18% FRL, 21% Minority, 88% Prof 2011
- E/la

<table>
<thead>
<tr>
<th>Group</th>
<th>% Prof 2011</th>
<th>% Prof 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Ed</td>
<td>53.8</td>
<td>73.3</td>
</tr>
<tr>
<td>Gen Ed</td>
<td>95.9</td>
<td>97.9</td>
</tr>
<tr>
<td>FRL</td>
<td>75.9</td>
<td>91.7</td>
</tr>
<tr>
<td>Paid</td>
<td>96.1</td>
<td>97.0</td>
</tr>
<tr>
<td>Bottom 25%</td>
<td>82.6</td>
<td>91.3</td>
</tr>
<tr>
<td>Top 75%</td>
<td>94.4</td>
<td>96.7</td>
</tr>
</tbody>
</table>
Results School L

- 682 students 87% FRL, 25% Minority, 58% Prof 2011
  - Mathematics

<table>
<thead>
<tr>
<th>Group</th>
<th>% Prof 2011</th>
<th>% Prof 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>62.1</td>
<td>84.0</td>
</tr>
<tr>
<td>White</td>
<td>72.7</td>
<td>71.9</td>
</tr>
<tr>
<td>Special Ed</td>
<td>34.4</td>
<td>57.6</td>
</tr>
<tr>
<td>Gen Ed</td>
<td>70.4</td>
<td>75.2</td>
</tr>
<tr>
<td>FRL</td>
<td>72.4</td>
<td>73.1</td>
</tr>
<tr>
<td>Paid</td>
<td>70.7</td>
<td>77.5</td>
</tr>
<tr>
<td>EL</td>
<td>52.4</td>
<td>81.1</td>
</tr>
<tr>
<td>Non-EL</td>
<td>73.6</td>
<td>72.8</td>
</tr>
<tr>
<td>Bottom 25%</td>
<td>38.6</td>
<td>46.3</td>
</tr>
</tbody>
</table>
GEORGIA’S COLLEGE AND CAREER READY PERFORMANCE INDEX (CCRPI)

Accountability is about giving leaders a roadmap for improvement. It is not about threatening schools. It gives them levers to effect change and a yearly opportunity to measure the pace of change. Every stakeholder in the school has access to data and can play a role in a continually improving success story.
About CCRPI

• Designed around a comprehensive definition of college and career readiness
• Scored on a scale of 0-100 points
CCRPI Main Components

- Overall score (0-100) comprised of:
  - Achievement (70%)
    - Content Mastery (40%)
    - Post E/M/High School Readiness (30%)
    - Graduate Rate (or Predictor for HS Grad Rate) (30%)
  - Progress (Student Growth) (15%)
  - Achievement Gap (15%)
  - Challenge Points (max of 10 bonus points)
    - ED/EL/SWD Performance
    - Exceeding the Bar Indicators
## CCRPI Achievement

### Content Mastery (End of Course Tests)
1. Percent of students scoring at Meets or Exceeds on the Ninth Grade Literature End of Course Test (required participation rate ≥ 95%)
2. Percent of students scoring at Meets or Exceeds on the American Literature End of Course Test (required participation rate ≥ 95%)
3. Percent of students scoring at Meets or Exceeds on the Mathematics I/GPS Algebra (transitioning to CCGPS Coordinate Algebra) End of Course Test (required participation rate ≥ 95%)
4. Percent of students scoring at Meets or Exceeds on the Mathematics II/GPS Geometry (transitioning to CCGPS Analytic Geometry) End of Course Test (required participation rate ≥ 95%)
5. Percent of students scoring at Meets or Exceeds on the Physical Science End of Course Test (required participation rate ≥ 95%)
6. Percent of students scoring at Meets or Exceeds on the Biology End of Course Test (required participation rate ≥ 95%)
7. Percent of students scoring at Meets or Exceeds on the US History End of Course Test (required participation rate ≥ 95%)
8. Percent of students scoring at Meets or Exceeds on the Economics End of Course Test (required participation rate ≥ 95%)

### Post High School Readiness
9. Percent of graduates completing a CTAE pathway, or an advanced academic pathway, or a fine arts pathway, or a world language pathway within their program of study
10. Percent of CTAE Pathway Completers earning a national industry recognized credential, or a passing score on a GaDOE recognized end of pathway assessment (operational in 2014-2015)
11. Percent of graduates entering TCSG/USG not requiring remediation or learning support courses; or scoring at least 22 out of 36 on the composite ACT; or scoring at least 1550 out of 2400 on the combined SAT; or scoring 3 or higher on two or more AP exams; or scoring 4 or higher on two or more IB exams
12. Percent of graduates earning high school credit(s) for accelerated enrollment via ACCEL, Dual HOPE Grant, Move On When Ready, Early College, Gateway to College, Advanced Placement courses, or International Baccalaureate courses
13. Percent of graduates earning 2 or more high school credits in the same world language (operational in 2013-2014)
14. Percent of students scoring at Meets or Exceeds on the Georgia High School Writing Test
15. Percent of students achieving a Lexile measure greater than or equal to 1275 on the American Literature EOCT
16. Percent of EOCT assessments scoring at the Exceeds level
17. Student Attendance Rate (%)

### Graduation Rate
18. 4-Year Cohort Graduation Rate (%)
19. 5-Year Extended Cohort Graduation Rate (%)
Progress

• Utilizes Student Growth Percentiles (SGPs)
  • Describe a student’s growth relative to academically-similar students
  • Percentiles range from 1-99
• Worth 15% of overall index score (15 points)
• Calculates the percentage of students demonstrating typical or high growth (≥35th percentile) across all subject areas
• That percentage is translated to a portion of the 15 points possible
Achievement Gap

• Compares mean z score performance of school’s lowest quartile with state average

• For each content area, schools’ receive a rubric-based score (1-4) based on gap size and gap change
  • Gap size: difference between mean z score of school’s lowest quartile and state average
  • Gap change: change in gap size from previous to current year

• Final content area score is higher of gap size and gap change

• Calculates percentage of points earned and translates to a portion of the 15 points possible
Challenge Points

- Opportunity for schools to earn additional points – maximum of 10
- Two components
  - ED/EL/SWD Performance
    - Acknowledges the academic performance challenge of having a significant number of ED/EL/SWD students
    - Utilizes performance flags which are triggered by performance targets
  - Exceeding the Bar Indicators
Additional Components

• Business rules
  • Minimum N = 15
  • Full Academic Year (FAY)
    • 65% enrollment
    • FAY presented a new challenge since EOCTs are now incorporated into accountability

• Performance Flags
  • State and subgroup performance targets
• Financial Efficiency Rating
• School Climate Rating
Considerations

• Achievement Indicators
• Weights
• Bonus points
• Minimum sample size
• Full academic year (FAY)
• Scoring/Benchmarking
Including Growth in Utah’s Comprehensive Accountability System

Judy W. Park, Ed.D.
Associate Superintendent
Utah State Office of Education
Principles

• Meeting standards (proficiency) and improving academic achievement (growth) are BOTH valued.

• All schools, including those that serve traditionally low performing students, should have an opportunity to demonstrate success.

• The system should include strong incentives for schools to improve achievement for the lowest performing students.

• Growth expectations for non-proficient students should be linked to attaining proficiency.

• Growth expectations for all students, including students above proficiency, should be appropriately challenging and meaningful.
Overview of UCAS

- The system consists of a total of 600 points that are split between Achievement and Growth.
- Structure and total points associated with each indicator:

  **Elementary and Middle Schools**
  - **Overall School Score**: 600 Total Points
    - **Growth**: 300 total points
    - **Achievement**: 300 total points
      - **All Students**: 200 total points
      - **Below Proficient Students**: 100 total points
      - **Percent at or above proficient**: 300 points

  **High Schools**
  - **Overall School Score**: 600 total points
    - **Growth**: 300 total points
    - **Achievement**: 300 total points
      - **All students**: 200 total points
      - **Below Proficient Students**: 100 total points
      - **Percent at or above proficient**: 150 points
    - **Readiness**: Graduation rate 150 points (HS only)
Growth – 300 Points

• Student Growth Percentiles - SGP

• Individual Students

• Aggregated at school level
  • All Students
  • Below Proficient Students
Incorporating Growth in UCAS – 2012-2013

• Assigns UCAS growth points (max total = 300 points) based on median SGP achieved in each content area by all students (AS) and below proficient (BP) students
• Growth points are averaged across content areas for each group
• Sum of points earned by both groups are reported for each school
Growth 2012-2013

- Rubrics used to assign growth points to AS and BP groups

<table>
<thead>
<tr>
<th>MGP Range</th>
<th>Growth Points Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-34</td>
<td>50</td>
</tr>
<tr>
<td>35-49</td>
<td>100</td>
</tr>
<tr>
<td>50-59</td>
<td>150</td>
</tr>
<tr>
<td>60 and above</td>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MGP Range</th>
<th>Growth Points Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-34</td>
<td>25</td>
</tr>
<tr>
<td>35-49</td>
<td>50</td>
</tr>
<tr>
<td>50-59</td>
<td>75</td>
</tr>
<tr>
<td>60 and above</td>
<td>100</td>
</tr>
</tbody>
</table>
Incorporating Growth 2013-2014

• Following the 2012-13 implementation, the approach was changed to address feedback from the policy advisory committee (PAC)
  • PAC expressed concern that collapsing MGP into 4 categories produced a ‘coarse’ indicator
  • Also, minor differences near the threshold can substantially impact outcome

• The new approach is based on the simple idea of multiplying MGPs for AS by 2 and add the MGP for BP
UCAS Growth Example 2013-14

• School A earns a growth score of 190 points and School B earns a growth score of 122 points.

<table>
<thead>
<tr>
<th>School</th>
<th>MGP of AS Group</th>
<th>MGP of BP Group</th>
<th>AS Points (multiply by 2)</th>
<th>Total Growth Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>59</td>
<td>72</td>
<td>118</td>
<td>190</td>
</tr>
<tr>
<td>School B</td>
<td>45</td>
<td>32</td>
<td>90</td>
<td>122</td>
</tr>
</tbody>
</table>
‘Scaling’ Growth Scores

• A limitation of this approach is that MGPs fall in a restricted range.
• The actual distribution is more compressed and doesn’t cover the full range of scores.
• As a result, scores are scaled to cover the full distribution
  • Linear transformation such that MGP of 30 receives lowest growth score and MGP of 70 receives highest growth score
UCAS Growth

• Assigns UCAS growth points (max total = 300 points) based on median SGP achieved in each content area by all students (AS) and below proficient (BP) students
• Growth points are averaged across content areas for each group
• Sum of points earned by both groups are reported for each school
Growth by Status - UCAS

2012 UCAS growth score (x) by prior year status (y), using UCAS 2011 status points.

Elementary/ Middle Schools  
(r=.266)  

High Schools  
(r=.138)
UCAS Growth Point Distribution by FRL Quartile

Elementary and Middle Schools

High Schools

Quartile 1 (top panel) = lowest poverty; Quartile 4 (bottom panel) = highest poverty
Growth for ELL Students - ELA

Each histogram shows the distribution of school MGP

The rows indicate ELL status. The top row is Not ELL and the bottom row is ELL only.

The columns indicate proficiency status. The left column shows results for below proficient, the right side shows proficient only.
OPPORTUNITIES AND THREATS MONITORING AND EVALUATION
Consolidated Subgroup

• Opportunities
  • Includes more students in a separate analysis who would have only been included in the “all students” group previously
  • Reduces the number of possible “hurdles” in an conjunctive model
  • Minimizes over-identification due to membership in multiple student groups
  • All lead to increased reliability of decisions

• Threats
  • Obscures small subgroup performance
  • Provides less information on who needs targeted assistance
Monitoring and evaluating the use of consolidated subgroups

• Monitor
  • Accuracy of classifications and calculations
  • Any negative consequences in treatment of low-frequency student groups
    • Scores stay low or go down
    • Not provided needed interventions

• Evaluate
  • To what degree is the double-counting of students in multiple subgroups contributing to a school being labeled as “needs improvement”?
  • Do consolidated subgroups include more students in the subgroup analysis than using traditional ESEA subgroups with a minimum n-size of 10?
  • Would the schools identified as needs improvement change using different methods of consolidation?
    • combine all at-risk subgroups
    • combine all non-white racial/ethnic groups but leave the others as standalone
    • combine ELL with SWD but leave the others as standalone
    • focus on the lowest performing 25%
    • focus on the lowest performing 33%.
Achievement Gaps

- Opportunities
  - Permits direct evaluation of inequity regardless of absolute performance
  - Facilitates comparison of schools with similar demographics but different achievement gaps

- Threats related to:
  - Metric (e.g. percent proficient, scale score)
  - Comparison Groups (e.g. demographic groups, performance based)
What is the Intended Outcome?

Monitoring and Evaluation: Gap Size

• Ensure schools with low achieving focal groups are not advantaged due to low performing or regressing reference groups
• There should not be a strong positive correlation between gap size and performance of the reference group
• Evaluate ‘Common Reference Threshold’ to ensure:
  • It doesn’t function as purely a status measure
  • Large gaps above the threshold are undetected
Monitoring and Evaluation: Gap Progress

• Example:
  • School A has a gap of 5 points in year one that was reduced to 3 points in year two. This 2 point reduction indicates a 40% improvement.
  • School B has a gap of 60 points in year one that was reduced to 36 points in year two. This 24 point reduction indicates a 40% improvement.

*Magnitude of progress can be distorted by the size of the starting gap*
Growth

• Opportunities
  • Proficiency alone is coarse and can be misleading
  • Better detect gains for students well below standard

• Threats related to:
  • Setting meaningful growth expectations
  • Dealing with model error
Monitoring and Evaluation

- Results are reasonably consistent or stable over time
- Results are not strongly correlated with variables not intended to be related to academic progress (e.g. school size, poverty)
- Growth thresholds are appropriate
  - Are target growth rates attainable?
  - Do students growing at these rates have a high probability of attaining proficiency/ readiness in a reasonable time frame?
Combining Measures & Producing Outcomes

• How will measures be combined to produce an overall result?
  • e.g. conjunctive, compensatory, profiles
• How will performance expectations be established within and/or across indicators?
  • Norm-referenced (e.g. ranking)
  • Criterion referenced (e.g. standard setting)
• How will results be communicated?
  • Type of summary score (e.g. letter grade, composite score)
  • Unit of information (e.g. indicator, subgroup, content area, combination)
Monitoring and Evaluation of the Overall System

- Includes many of the approaches discussed previously, but at the total score level
- Include impact of sampling error
- Examine discrepancy between/among indicators to identify ‘problematic profiles’
- Track improvement over time, particularly for priority and focus schools
Questions/ Discussion