Accountability as Inquiry

Brian A McNulty Ph. D.

Explore the ESEA reauthorization and upcoming changes
Challenge current policy solutions
Present some emerging ideas and opportunities
Celebrate progress in OIP and OLAC

Campbell’s Law

“When any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

Campbell 1976

• When quantitative measures such as test scores are used to make key decisions:
  • The measures themselves are subject to corruption pressures and,
  • In addition, the high-stakes distort and corrupt teaching and student learning

Welner (2013).

Campbell’s Law

• Corollaries:
  • Classroom instruction leading up to the standardized test is now preparation for the teacher’s job evaluation
  • Preparation for the test becomes the job itself!

Welner (2013).

ESEA/NCLB Reauthorization – Eliminated

• AYP
• Universal proficiency, outcomes, and dates
• “Cascade of sanctions”
  – Specific interventions of low performing schools

ESEA/NCLB Reauthorization – Eliminated

• Title I portability
• Highly qualified teachers
  – Teacher evaluation
• RttT
• Reading First
ESEA/NCLB Reauthorization - Every Student Succeeds Act (ESEA) - Continued

- Challenging standards
- States still have to test student in math and reading in grades 3-8 and once in high school and still report by subgroups
- No role in evaluation of teachers from USDOE

State participation in NAEP

PARRC ELA

<table>
<thead>
<tr>
<th>PERFORMANCE LEVEL</th>
<th>Public Districts</th>
<th>Community Schools</th>
<th>Total Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>Proficient or Above</td>
<td>Advanced</td>
<td>7,870</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>Accelerated</td>
<td>43,754</td>
<td>36.5</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>51,807</td>
<td>26.5</td>
</tr>
<tr>
<td>Below Proficient</td>
<td>Basic</td>
<td>20,989</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>15,730</td>
<td>13.1</td>
</tr>
</tbody>
</table>

PARRC Math

<table>
<thead>
<tr>
<th>PERFORMANCE LEVEL</th>
<th>Public Districts</th>
<th>Community Schools</th>
<th>Total Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>Proficient or Above</td>
<td>Advanced</td>
<td>861</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Accelerated</td>
<td>22,692</td>
<td>26.9</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>24,374</td>
<td>28.8</td>
</tr>
<tr>
<td>Below Proficient</td>
<td>Basic</td>
<td>21,912</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>17,959</td>
<td>20.4</td>
</tr>
</tbody>
</table>

PARRC Algebra I

<table>
<thead>
<tr>
<th>PERFORMANCE LEVEL</th>
<th>Public Districts</th>
<th>Community Schools</th>
<th>Total Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>Proficient or Above</td>
<td>Advanced</td>
<td>2,399</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>Accelerated</td>
<td>39,008</td>
<td>35.6</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>31,821</td>
<td>29.3</td>
</tr>
<tr>
<td>Below Proficient</td>
<td>Basic</td>
<td>24,477</td>
<td>22.5</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>10,993</td>
<td>10.1</td>
</tr>
</tbody>
</table>
ESEA/NCLB Reauthorization
• 7% set-aside for turnarounds
• Some states can apply to use local tests-experiment with new types of assessment.
• Only 1% of special ed. students can get alternative test
• Consolidation of programs

What’s NEW: What Does it all Mean
• SEA’s will have more discretion over:
  • Their accountability systems
  • Their evaluation systems
  • Interventions in low performing schools

Waiver from Ohio’s State Tests available
• Districts and schools have the opportunity to submit a proposal to explore alternative approaches to state tests in 2016-2017.
• Ohio law allows the state superintendent of public instruction to approve up to 10 waivers.
• A waiver permits the district or school to develop and use an alternative assessment system for the district’s students in place of the state assessments required under law.

Assessment

The amount of new technical knowledge is doubling every year

Darling-Hammond & Adamson 2014
Schools must teach disciplinary knowledge in ways that also help students learn how to learn, so that they can use knowledge in new situations (and new ways.)

The skills in greatest demand are the non-routine interactive skills that require collaborative invention and problem solving.

There is a need to measure both core knowledge and higher order thinking skills.

The return of performance assessments

- During the 1990’s many states developed assessment systems that included locally administered performance assessments (CT, KY, ME, MD, NE, NH, NJ, NY, OR, VT, RI, WA, WI, WY and others).

Research found that when teachers become experienced in developing and evaluating high-quality performance assessments, they are better able to design and deliver higher-quality learning experiences.

The return of performance assessments

- Research done then found that these performance assessments improved the quality of instruction.
- However most states discontinued these with the adoption of NCLB requirements of annual grade level testing.
The Gordon Commission on Future Assessment in Education

“To be helpful in achieving the learning goals laid out in the Common Core, assessments must fully represent the competencies... The best assessments can accelerate the acquisition of these competencies if they guide the actions of teachers and enable students to gauge their progress.

Darling-Hammond & Adamson 2014

Two alternatives are particularly worthy of consideration:

1. Combining multiple measures that include inputs as well as outputs; and
2. Inspectorate systems incorporating self-evaluations coupled with site visits conducted by disinterested but qualified visitors representing the state or an accreditation group.

Mathis 2015

Multiple measures -considerations

1. Adequate student opportunities and resources to achieve each state's goals;
2. Continued development of multiple-measure approaches that strive for balance and clarity;
3. Cautious use of standardized test scores;
4. Avoidance of data aggregation into a single score;

Mathis 2015
Unfortunately, “multiple measures” is an elastic term that includes an eclectic variety of elements. The term can mean many different things and thus result in many different policy approaches.

Mathis 2015

Advocates of multiple measures often speak of a “dashboard” of decision data. But their validity, as a measure of school quality, is still open to question.

Mathis 2015

“Multiple measures” has served as a bridging concept between different policy camps. Darling-Hammond and Hill, for example, released companion reports addressing elements to be included in the next generation of school evaluation systems.

Mathis 2015

2. School Self-Evaluations Plus Inspectorates

Self-evaluation combined with inspectorate systems continue to be the norm in most developed countries (OCED).

Mathis 2015

Inspectors

1. Development and implementation of school visitation teams, with a priority on higher need schools;
2. External reviews focusing on guidance and support rather than sanctions;
3. Trained and qualified reviewers who meet prescribed standards; and
4. Multiple stakeholders involved in the design of state’s evaluation/inspectorate program.

Mathis 2015

The self-evaluation report becomes a foundational document for the inspection team.

- Through interviews and data review, the team seeks to verify such things as
  - student expectations,
  - the comprehensiveness of assessment,
  - curricular adequacy,
  - professional development, and
  - available supports and interventions.

Mathis 2015
No evaluation system by itself is capable of overcoming the deficiencies of a school or community lacking resources.

Mathis 2015

The history of educational reform has been one of consistent failure as the field moves from fad to fad. Over time policy makers have repeatedly implemented simple solutions for reform based upon little if any research.

Good & Lavigne (2014)

AERA Issues Statement on the Use of Value-Added Models in Evaluation of Educators and Educator Preparation Programs (11.11.15)

The statement, cautions against the use of VAM for high-stakes decisions regarding educators.

The AERA statement speaks to the formidable statistical and methodological issues involved in isolating either the effects of educators or teacher preparation programs from a complex set of factors that shape student performance.
Multiple researchers have analyzed the VAM model (and) found that estimates produced by the model are significantly biased.

Table 1. Instability of Value-Added Teacher Rankings in San Diego and 5 Florida Counties

<table>
<thead>
<tr>
<th>Locale</th>
<th>Bottom 20% in Year t</th>
<th>Top 20% in Year t</th>
<th>Bottom 80% in Year t+1</th>
<th>Top 80% in Year t+1</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Diego, CA</td>
<td>65%</td>
<td>71%</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Duke County, FL</td>
<td>70%</td>
<td>67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dural County, FL</td>
<td>67%</td>
<td>61%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillsborough County, FL</td>
<td>67%</td>
<td>67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange County, FL</td>
<td>59%</td>
<td>65%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palm Beach County, FL</td>
<td>60%</td>
<td>68%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

American Statistical Association Statement


• “VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward student outcomes.”

As classrooms vary, so will the value-added scores obtained by teachers.

When three years of rankings are used,

• reliability is even worse
• 74% of teachers ranked in the bottom quintile shift out of that quintile
• 56% of teachers ranked in the top quintile shift out of that quintile
The only way to know if measures of teaching truly identify effective teaching and not some unmeasured student characteristics is by randomly assigning teachers to students

Gates Foundation, (2013)

VAM lacks validity for the purpose of high-stakes decisions regarding individual teachers

Yeh, (2013)

Because of this (variability), you simply can-not ethically and legally use these data for high-stakes decisions about teacher competence,

However, these data might be used for purposes of formative evaluation of teachers.

Berliner 2014

What does and doesn’t work!

What does and doesn’t work..

• Does the size of the school matter?
• Reducing size doesn’t improve outcomes

William, 2011

What does and does not work..

Does paying teacher bonuses work?
The scores of the students taught by teachers who were offered the bonuses were no higher than those taught by other teachers

William, 2011
What does and does not work..

Reconstitution
- A comparison of similarly low-performing schools that were not reconstituted shows that they improve at the same rate.

What does and does not work..

Charters
- 1/3 of charter schools get worse results, and
- 1/6 get better results.
- The net effect of the introduction of charter schools, therefore, appears to be a slight lowering of student achievement.

What does and does not work..

More than 80% of charter schools are either no better or worse than traditional public schools at securing math and reading gains for students. The states most associated with extensive charter school laws, parental choice, and accountability systems (Florida, Ohio, and Texas) rank highest among the states where traditional public school students outperform charter school students.

What does and does not work..

- Teachers’ aides actually lowered the performance of students they were assigned to help.
- Textbooks there was little evidence that changes in textbooks alone had much impact on student achievement.
- It was only when schools changed teaching practices and student interactions that a significant impact on achievement was found.

Teacher unions are the problem

Only five states do not allow collective bargaining for educators, effectively banning teachers unions. Those states and their SAT/ACT rankings are as follows:
- South Carolina – 50th
- North Carolina – 49th
- Georgia – 48th
- Texas – 47th
- Virginia – 44th

States that are more heavily unionized tend to have higher achievement scores (Strauss, 2010).
What does work

WHAT WORKS

Framework- Key Practices

1. Use data well
2. Focus and limit goals and strategies
3. Select and implement shared instructional practices
4. Implement deeply
5. Monitor, provide feedback and support
6. Collaboratively inquire and learn
What is needed

• Universal Preschool
• Family wrap around services
• Better training for teachers (pre-service and in-service)

In the breakout session

Our 3 most promising strategies are not working well

• Classroom observations
• Professional learning communities (TBTs) and
• Coaching
• What to do about this

The Mirage Study Findings

School systems are not helping teachers understand how to improve—or even that they have room to improve at all.

Other work across Ohio

New modules & Videos

• Diverse Learner
• Effective Curriculum Practices
• Ability Grouping