

On Accountability, Evaluation and Systemic Change

Henry Braun, Ph.D.

Boisi Professor of Education and Public Policy

July 22, 2018



Why Accountability?

- Why hold public services accountable?
 - Democratic function
 - Research/improvement function
- What are the main components of accountability?
 - Evidence-based evaluation (monitoring)
 - Feedback (for improvement)
 - Consequences
- What is the track record of *accountability* as an instrument of *improvement*?
 - For high-stakes accountability → Generally poor
 - For evaluation alone → Mixed



Accountability vs. Improvement: Dynamic Tensions

- Accountability (with consequences) triggers Campbell's Law

The more 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.

- Even relatively weak accountability can induce unintended negative consequences
- If monitoring systems are to support continuous improvement, then the data generated should be
 - Comprehensive
 - Accurate
 - Relevant
 - Timely



The Need for Improvement

In the U.S. and many other OECD countries, children born to disadvantage face:

Gates ... to opportunity

Gaps ... in the accumulation of *human capital* (HC) and *social capital* (SC)

Gradients ... in the relationships of life outcomes to HC and SC

(along with a tighter coupling of HC and SC))

At the same time, we all confront:

Increasing inequality at home

Increasing competition abroad

The challenge lies in designing, implementing and sustaining scalable changes in education that make a positive difference in the life chances of all children.



Considerations in Data Use for Improvement (Hargreaves & Braun)

- The ***nature of the data*** in type, quality and range;
- The ***indicators*** of growth, of progress towards higher standards and threshold targets, and of benchmarked comparisons with peers that are derived from the data;
- The ***interface and interaction*** between the data dynamics of accountability and improvement systems respectively;
- The ***consequences*** attached to high and low performance; and
- The ***culture and context*** of data use.



Further Considerations for *Data-Informed* Improvement

- ***Measure what is valued instead of valuing only what can easily be measured.*** Metrics and indicators should accurately reflect the range and levels of the learning goals and other priorities set by the state such as critical reasoning, emotional and social learning, creativity and teamwork.
- ***Create a balanced scorecard.*** Collect evidence on a regular schedule from different sources to capture different aspects of system functioning and multiple student outcomes. The student data and administrative data that are routinely collected and reported in most school systems do not render a sufficiently complete picture of the education that students receive, nor of the factors that affect students' learning.



- **Articulate and integrate the components of the DDIA system both internally and externally.** Internally, different data types (e.g. formative, interim and summative assessments) and their use should complement rather than contradict one another. For this reason, all assessments should be coherent with a common set of content and performance standards. Externally, DDIA should cohere with other parts of the improvement and accountability system.
- **Insist on high quality data.** Institute a regular and rigorous quality assurance audit of all indicators used for improvement and accountability. In particular, test-based indicators used for high stakes decisions should meet industry standards with respect to accuracy, reliability, year-to-year stability, and validity.



- ***Test prudently, not profligately.*** One of the objections to increasing the level of sophistication of tests and indicators is the increased cost. But it is counterproductive to control costs by settling for lower test quality that impedes improvement, diminishes authentic accountability, and undermines the system's credibility. A widely used and successful alternative is to reduce the scope and frequency of testing.
- ***Establish improvement cultures of high expectations and high support.*** Set challenging performance standards for students and attainable benchmarks for schools, with the proviso that adequate support for continuous school improvement will be provided by the system



- ***Move from thresholds to growth.*** Systems should limit the use of imposed numerical targets tied to threshold criteria as these induce a host of perverse incentives. Indicators based on student progress, by comparison, encourage educators to address the needs of all students and to keep moving forward without the anxiety about reaching one particular target at a specified time.
- ***Narrow the gap to raise the bar.*** Test score gaps reflect, in large part, differences in family and community assistance available to students, as well as differences in the levels of resources and capacity within schools and school districts. Evidence of such disparities should trigger support for both students and schools.



Linking Monitoring Systems to Improvement Models

- *Assign shared decision-making authority, as well as responsibility for implementation, to strong professional learning communities.* The DDIA system should support high trust professional communities characterized by collective responsibility for all students' success, and in which data-informed discussions are valued alongside other effective modes of professional collaboration. Mutual support among educators then becomes the norm and students are less likely to “fall through the cracks” as they move from one class to another. High-trust environments assign significant authority to professional communities for shared decision-making in relation to **data-informed judgments**.



Ontario: Essential for Some, Good for All

➤ Ontario

- Population > 14 million (highly diverse)
- Area > 1 million km²
- 3 school sectors: Public (English), Catholic (English), Franco-Ontarian
- Education for All (White paper → Education policy initiatives)
- EQAO testing system (Reading, Writing & Math: selected grades)
- Threshold benchmarks for schools
- Weak accountability (relative to U.S.)

➤ ESGA

- 3 year project, \$57M
- Focus on
 - Improving performance of students with special needs
 - Reducing special education identification rates



ESGA: Organization

- Ministry transferred responsibility to the Council of Ontario Directors of Education (CODE)
- CODE
 - Developed proposal template and instructions (\$50k/board initially)
 - Reviewed proposals and provided feedback
 - Sent expert educators to work with school boards
 - Approved final proposals and disbursed funds
 - Little Ministry involvement
- Project process
 - Maintained ongoing communication with boards
 - Annual convening of all boards
 - Review and feedback on annual reports



ESGA: Evaluation (Hargreaves & Braun)

- 10 Boards representing range of 72 boards
- 3-day site visits by BC teams
- Interviews with top Ministry officials
- Data sources
 - On-site observations and interviews
 - Web-survey
 - EQAO results
 - Document review



ESGA Architecture

- ***Inspiring Beliefs*** that motivate widespread participation;
- ***Leading from the Middle*** by a respected third party of former superintendents who were actively supported by a large majority of their provincial colleagues;
- ***Local Authority*** and flexibility that allows and insists on responsiveness to the diversity of local needs and circumstances;
- ***Local buy-in*** → Investment of professional capital to assure success
- ***Collective Responsibility*** for all students' learning at the school and school board levels - especially between special education staff and their colleagues with curriculum and classroom responsibilities; and
- ***Intensive Interaction*** that connects everyone and creates coherence among all policy elements by constant monitoring, mentoring and cross-pollination of insights, ideas and activities.



ESGA Outcomes (I)

- Improvement in EQAO results (Reading, WRITING) – Closing gaps
- Reduced rate of increase in ID rates
- More mainstreaming of Spec Ed students
- Greater professional collaboration and increase in collective responsibility for student development (school-level)
- Increased use of tiered interventions, differentiated instruction and assistive technologies
- Changes in assessment practices and data use
- Building capacity to address needs of at-risk students



ESGA Outcomes (II)

- Positive changes in working relationships in Boards' central offices
- Incremental changes vs. Radical restructuring
- Benefits of local buy-in offset loss of central control
- Pilot efforts → Long-term sustained changes
- Subsequent reforms capitalized on
 - Successful boards as regional leaders in next phases
 - Local change agents
 - Momentum of ESGA successes



Teacher Professional Development (TPD)

Multiple flavors

- One-offs (e.g. 3 hour presentation on classroom management)
- Extended support (e.g. Initiating and sustaining a *Professional Learning Community*)
- Link to a general program (e.g. Cognitively Guided Instruction; Introduction to social-emotional learning and how to bring it into your classroom)
 - Teacher knowledge and/or practices
- Embedded in a larger, targeted reform effort
 - Public (e.g. Success for All)
 - Commercial (e.g. Curriculum Associates- iReady Diagnostic Assessments)



TPD: Evaluation (I)

Many frameworks for TPD evaluation (e.g. Guskey)

- 1. Participants' reactions
- 2. Participants' learning
- 3. Organization support and change
- 4. Participants' use of new knowledge and skills
- 5. Student learning outcomes

Multiple data sources needed to document results at each level.



TPD: Evaluation (II)

- Most (short) TPD programs are not evaluated or only at Level 1
- Longer programs may be evaluated up to Level 4
- More extensive evaluations usually done under the auspices of specific interventions, with TPD only one component of the intervention
- Appears to be rare that variations in TPD (quality, extent) are linked to differences in student outcomes
- More typical that variations in TPD included in measures of *Fidelity of Implementation*



TPD: Best Practices

- offered for a **longer duration and greater frequency**
- involves teachers directly for more hours in **active, engaged learning** activities and environments
- **focused on a particular content area**, such as geometry or astrophysics, and allows teachers to gain knowledge on how to teach the content to their students
- **coherent to teachers' needs and circumstances**
- involves teachers learning from their peers through **collective participation**



TPD Evaluation: Challenges

- Need for longitudinal data to track changes in
- Difficulties in linking pedagogical practices to student outcomes
- Variation in TPD quality and/or uptake correlated with variation in other relevant factors

Example: Intensive Partnerships for Effective Teaching

- Funded by Gates Foundation and Partner sites (\$575M)
- RAND/AIR evaluation of 5-year implementation: Found little evidence of impact on
 - Student learning
 - Graduation rates



TPD: Conclusions

- TPD should be one component of a “reform package” that has specific targets for student learning
- Reform/TPD should be grounded in an inspiring/motivating vision of instruction
- Reform should be adaptable to local contexts without sacrificing key drivers
- Comprehensive data collection/analysis infrastructure should be in place for both real-time monitoring, provision of feedback, and summative assessment
- “Means conceal more than they reveal” →
Pay attention to variation, especially at the extremes!





Thank You

