Multiple Indicators of School Effectiveness

"It is time for states to move beyond NCLB's myopic focus on two test scores to look more broadly at the resources, processes, and outcomes at individual schools. Working with stakeholders, states should develop valid, reliable, and fair measures of schools that lead to transparency, support and improvement, rather than the unfair labels and punishments we have seen as a result of NCLB."

— NEA President Dennis Van Roekel

Multiple indicators of school effectiveness have potential benefits in the area of school transparency, self-improvement and accountability. However, any discussion of measures of school effectiveness should begin with a cautionary tale drawn from the recent history of federal and state efforts to measure school performance under No Child Left Behind.

State efforts to monitor standards-based school effectiveness began in the 1980s, accelerated under the Improving America's Schools Act of 1994 (IASA), and became subject to the detailed accountability system prescribed by the No Child Left Behind Act (NCLB) in 2001.[1] NCLB's accountability system was unfortunately based on an under-resourced goal of 100 percent proficiency by 2014, a narrow set of standardized tests, and a system of labels and punishments for schools that missed any one of 37 Adequate Yearly Progress (AYP) categories. Although the law created greater transparency about the relative performance of student subgroups, its focus on narrow high-stakes tests was often counterproductive, leading to: teaching to the test; a narrowing of the curriculum;

focus on students on the cusp of proficiency at the expense of others; push out of students who might produce poor test results; a focus on lower-order thinking skills; exodus from the teaching profession by teachers discouraged by the rote drilling of students; and the indiscriminate labeling of a majority of public schools as failures.[2] As a result of NCLB's failures, the federal government is now granting waivers from many of the requirements of NCLB.[3]

The unintended consequences of NCLB strongly suggest that great care should be taken in constructing any system of school measurements and accountability, beginning with hard questions about the purposes and use of the system. One set of such questions was posed in a Council of Chief State School Officers (CCSSO) study[4] on designing accountability systems published in 2002 as NCLB had only begun to narrow state systems:

- What are the purposes of the accountability system?

- What are the main contexts, political and otherwise?

- What are the main legal and policy constraints or specifications?

- What are the units of performance, accountability, and reporting?

- What are schools/students (or others) to be held accountable for?

- What accountability decisions will be made, and with what consequences?

- How will results be reported?

- What data are available and will be used in the accountability system?

- How will data be combined to make an accountability judgment?

- How will the accountability system be monitored and evaluated?

NCLB's narrow accountability system took attention and resources away from fundamental questions like these, but as states are given more latitude as NCLB falters, they continue to be of importance, including the fundamental issues of what gets measured and why.

The Rationale for a System of Indicators

A recent comprehensive Rand Corporation report titled "Expanded Measures of School Performance" (hereinafter Rand report) offered several potential reasons for having a school indicator system:

1. Monitoring to provide "periodic snapshots" of how the school is doing.

2. Diagnosis and prescription, e.g. allowing districts to identify areas where schools are struggling and guiding educators on how to improve school performance through, e.g. professional development or changes to the curriculum.

3. Signaling what is important by measuring it in the indicator system.

4. Accountability for rewards and sanctions.[5]

The Rand report warned that it is important to ensure that assessments validated for one of these four purposes (e.g. diagnosis) not be used for another purpose (e.g. accountability) without careful examination of its appropriateness. It also cautioned that attaching consequences to any measure "can lead to score corruption and undesirable narrowing of effort…at the expense of those that are not measured."[6]

With regard to an expanded set of indicators, the Rand report noted several possible benefits: greater validity through the comparison of overlapping dimensions; greater reliability; and a reduced risk of narrowing the focus of education.

NCLB's system of measurements largely served only the last two of the four purposes identified by the Rand report for an indicator system: signaling what is important (through its math and reading tests and grade span science tests) while inadvertently signaling that everything else was unimportant; and setting up a system of labels, rewards and sanctions. In doing so, NCLB missed out on a highly productive purpose—providing officials, educators and other stakeholders with transparent and timely information that can be used to diagnose problems, bring support and make changes where needed, and allow for continuous improvement.[7]

What Should be Measured

Another fundamental question about indicator systems is what to measure. NCLB's set of measures consisted of an extremely narrow of school outcomes (standardized test scores), which ignored both school inputs and processes. As such, NCLB provided little diagnostic direction to schools and educators and diverted attention from inputs and processes.

A broader set of measures was identified by a group of researchers working with educators in 1998 and provides a useful framework for thinking about state indicators. This framework sets out three areas: inputs, process, and outcomes. The researchers provided the following examples provided below:[8]

Imagine that the last ten years of public policy, government, and school action had been focused on reporting and improving a comprehensive list of important school inputs, processes, and outcomes, rather than focusing on one set out of outcomes, NCLB test scores, and punishments based on those outputs. A much deeper understanding of schools would have resulted, allowing for much better diagnosis and a broader template of solutions. Even in the area NCLB focused on—student learning outcomes—the focus was too narrow. For example, multiple measures of student learning that give educators immediate feedback on student learning and take into account growth are superior to a handful of once a year standardized tests.[9]

Some states have made efforts to collect and publish information on inputs and processes, including such information on report cards, but the notion that all stakeholders should have a role in improving those inputs and processes has taken a back seat to labeling schools based on test outcomes. According to the Rand report, some of the state measures included on report cards for information purposes only include: "student demographics, attendance, mobility…school per-pupil spending, course offerings, technology in the classroom, parental attendance at teacher conferences, community demographics, prime instructional time, the availability of arts programming, the number of first graders who attended full-day kindergarten, grade inflation, on-time progression through school, school principal turnover rate, the presence of a written student code of conduct, a crisis management plan, and availability of public school choice."[10] College readiness indicators were used in some states for informational purposes, in others for accountability purposes.[11]

INPUTS	PROCESSES	OUTCOMES
■ Financial (e.g., expenditures per average daily attendance; local, state, and federal revenue per average daily attendance)	■ Curriculum and instruction	■ Academic achievement (e.g., SAT, exit exam, writing tests)
■ Personnel (e.g., pupil-teacher ratio, average teacher salary)	■ Implementation of policies (e.g., admission, grading, promotion, etc.)	■ Accomplishments (e.g., graduation rates, college attendance)
■ Facilities	■ Diverse educational opportunities	■ Attitudes
■ Equipment	■ Parent involvement	■ Retention/dropout rates
■ Materials	■ Leadership (e.g., planning, style, efficiency, etc.)	■ School safety
■ School policy/law		■ Discipline
■ Student attributes (e.g., predominant socio-economic level, average ability)		■ College enrollment and completion rates

In the area of outcomes, states are now looking to refine how output data is measured and provide for more effective use of the data. For example, the CCSSO has a Roadmap for Next Generation Accountability Systems that highlights several aspects of such systems. According to CCSSO, these systems should clearly articulate state expectations for school and district performance so that all stakeholder actions are aligned toward ensuring students are ready for college and careers; differentiate the performance of schools and districts so that schools and districts receive appropriate support and interventions where needed; empower and engage educators and stakeholders through transparent, timely reporting of actionable data; and foster a commitment to innovation and continuous improvement.[12] Although CCSSO focuses almost solely on outputs, it insists that those outputs be usable for improvement, not simply for sanctions.

Tradeoffs in Choosing Measures

Although a wide range of informative indicators is available by which to measure schools, the Rand report describes some significant tradeoffs for those developing multiple indicator systems. In addition to closely examining the purpose of any new measures (monitoring, diagnosis, signaling, or accountability), the Rand report urges that the following should be taken into consideration: breadth versus focus; complexity versus transparency; comprehensiveness versus affordability; uniformity versus flexibility; formative versus summative; signaling versus preventing corruption of measures; and adjusting versus not adjusting for input.[13]

The report also raises technical questions that should be used with any new measures, including validity, reliability, and fairness.[14]

These tradeoffs and technical questions reinforce the cautionary tale that we began with. For a system of multiple indicators to work, we must learn from the mistakes of NCLB or else repeat them.

NEA Policy Recommendations

■ Education stakeholders in each state should consider ensuring that the state has a broad range of indicators of school supports, process, and outcomes. The system of indicators should be used to provide support for struggling schools and students; facilitate self-evaluation and improvement at the classroom, school, district and state levels; create transparency; and, where appropriate, promote accountability that helps, rather than punishes, schools.

■ Student learning can be one indicator of school effectiveness but should not be measured solely by student performance on one or two tests. Instead, multiple, valid, and reliable measures of student learning should be considered, beginning with a professional assessment by the classroom teacher. Such measures of student learning might include district- and school-level assessments; classroom-level written, oral, performance-based, or portfolio assessments; and written evaluations.

■ Valid measures of student learning should be developed with the agreement of teachers and other stakeholders and should take into account the multiple factors influencing a student's learning. Other indicators of school performance, which should also be identified in cooperation with education stakeholders, might include, for example, graduation rates, attendance rates, and the number and percentage of students participating in rigorous coursework.

■ School resources and processes should also be reported in a transparent way, for example funding, leadership and staff experience; class size (student-teacher ratio); number of National Board certified teachers; number of certified counselors, nurses, and other support staff per student; school building and environmental ratings; school anti-bullying policies; access to courses with 21st century skills; professional development and instructional improvement strategies; parent engagement programs; and access to

libraries, science laboratories, health care, nutritional meals, and before- and after-school programs.

- No indicator should ever be adopted without consensus or without careful attention to its use, validity, reliability, fairness, and any potential negative consequences.

- The broad indicators set out in NEA's Great Public Schools Criteria should inform discussions of school quality, rather than test-based accountability, These criteria include: quality programs and services that meet the full range of all children's needs so that they come to school every day ready and able to learn; high expectations and standards with a rigorous and comprehensive curriculum for all students; quality conditions for teaching and lifelong learning; a quali-fied, caring, diverse, and stable workforce; shared responsibility for appropriate school accountability by stakeholders at all levels; parental, family, and com-munity involvement and engagement; and adequate, equitable, and sustainable funding.[15]

RESOURCES[16]

Council of Chief State School Officers (CCSSO) (2011). *Roadmap for Next Generation State Accountability Systems (Working Draft)*, available at http://www.ccsso. org/documents/Roadmap.pdf.

Darling-Hammond, Linda. Evaluating 'No Child Left Behind', Nation, May 21, 2007.

National Center for Fair and Open Testing (Fairtest) (2012). *NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure?* Available at http:// fairtest.org/sites/default/files/NCLB_Report_ Final_Layout.pdf.

National Education Association (2006). *ESEA: It's Time for a Change! NEA's Positive Agenda for the ESEA Reauthori-zation.* Available at http://www.nea.org/assets/docs/HE/TM-NEAP osetiveAgendafortheESEAReauthorization.pdf.

National Education Association (2011), *NEA Policy Brief: Beyond Two Test Scores: Multiple Measures of Student Learning and School Accountability.* http://www.nea.org/assets/ docs/PB38beyondtwotestscores2011.pdf.

Schwartz, H. L., Hamilton, L. S., Stecher, B. M., & Steele, J. L. (2011). *Expanded Measures of School Performance.* Santa Monica, Calif.: The Rand Corporation, available at http://www. rand.org/content/dam/rand/pubs/technical_ reports/2011/RAND_TR968.pdf.

REFERENCES

[1] Council of Chief State School Officers (CCSSO) (2011), *Roadmap for Next Generation State Accountability Systems (Working Draft)*.

[2] See e.g. Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*, New York: Basic Books; National Center for Fair and Open Testing (Fairtest) (2012), *NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure*; Linda Darling-Hammond, Evaluating 'No Child Left Behind', Nation, May 21, 2007.

[3] Secretary of Education Arne Duncan (2012, January 6), *Escaping the Constraints of 'No Child Left Behind'*, Washington Post, The Department of Education waiver program continues the requirements of test-based accountability, as well as a focus on graduation rates, but allows for testing in subjects beyond reading and math and also requires that student growth be considered, There is also less emphasis on one-size-fits-all punishments for schools, See Department of Education (2011), ESEA Flexibility Principles, available at http://www.ed.gov/ esea/flexibility/documents/esea-flexibility-acc.doc.

[4] Gong, B., Blank, R., & Manise, J. (2002). *Designing School Accountability Systems: Towards a Framework and Process*. Washington, DC: CCSSO, available at http://programs.ccsso.org/content/pdfs/designing_school_acct_syst.pdf.

[5] Schwartz, H. L., Hamilton, L. S., Stecher, B. M., & Steele, J. L. (2011). *Expanded Measures of School Performance*. Santa Monica, Calif.: The Rand Corporation, 7.

[6] Schwartz et al., 8.

[7] A similar list of goals can be found in CCSSO (2011), 3.

[8] McLean, J., Snyder, S., Lawrence, F., *A School Accountability Model* (1998), Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, available at http://www.eric.ed.gov/PDFS/ED428440.pdf.

[9] National Education Association (2011), *Policy Brief: Beyond Two Test Scores: Multiple Measures of Student Learning and School Accountability*, Multiple measures of student learning should be one element of multiple measures/indicators of school effectiveness,

[10] Schwartz et al., 13. The Rand report's review of emerging input, process, and outcome measures found three main categories: "establishing a safe and supportive school environment; identifying students who are at risk of failing; and improving student outcomes." Schwartz et al., 15-24.

[11] Schwartz et al., 13.

[12] CCSSO (2011), 3.

[13] These tradeoffs are expanded upon detail at Schwartz et al., 29-31.

[14] Schwartz et al., 31-33.

[15] National Education Association (2006), *ESEA: It's Time for a Change! NEA's Positive Agenda for the ESEA Reauthorization*. NEA is currently in the process of developing an indicator system closely related to these criteria.

[16] The external resources on this list provide important background on the subject of multiple indicators, but do not necessarily reflect NEA policy on specific issues.

6   NEA Education Policy and Practice Department | Center for Great Public Schools | 1201 16th St., NW, Washington, D.C. 20036

2013 (PB45)