



SANDLER  
FOUNDATION

## Questions & Answers about Multiple Measures

---

**Q. Why should federal lawmakers encourage the use of multiple measures?** At the state level, there's a growing acknowledgement that standardized assessments in a handful of core subjects aren't enough to adequately measure school performance. Critics also have long raised concerns that NCLB overemphasizes math and English language, emphasizes only proficiency to the detriment of students well above and below the proficiency threshold on the test, allows states to set their own low bar for proficiency, and uses measures and metrics in too rigid a manner to determine Adequate Yearly Progress. There is growing interest among states and in the policy community to seize state momentum in using measures such as college- and career-readiness and dropout rates as key indicators of progress and to consider other factors that are associated with school performance and student success for a variety of purposes, including accountability, diagnostics, and targeting interventions.

**Q. How many states now evaluate schools using a broader range of measures than those required by NCLB?** According to RAND Education's new report, "Expanding Measures of School Performance," nearly half of all states—24 of 50—have developed their own accountability systems that supplement federal measures, and 20 publish information that includes additional indicators they feel are relevant for informing the public, implementing accountability measures, and developing a more sophisticated understanding of the contexts in which their schools operate. (*The report is now available at RAND.org.*) In a separate report, Achieve, Inc. identifies other indicators not addressed by RAND. For example, 12 states are using the percentage of high school graduates who earn a college- and career-ready diploma for a range of purposes, including school-level reporting/incentives, and two states (Texas and Louisiana) use the measure for accountability.

Generally speaking, these states are examining outcomes for student performance in *additional subjects* (most often history or social studies, but in some cases civics, economics, or even geography); *growth* in student performance over time; indices that take into account student achievement *along the entire spectrum of performance* rather than only a threshold of proficiency used for accountability purposes; and *measures of college- and career-readiness*, including SAT, ACT, and Advanced Placement test participation and scores, as well as drop-out and graduation rates.

Increasingly, states and districts are also using measures that take into account other factors that lead to increased student performance and school success. States are developing measures and metrics of safe and supportive learning environments, graduation-risk indicators, and results of "interim" academic assessments. Some are also experimenting with comparing the performance of schools with similar demographic characteristics, using school surveys and third-party inspections to glean more information, and examining more comprehensive measures of student transition to adult life, including employment rates and enrollment in college courses or job training.

---

- 
- Q. **Are these measures used for accountability only or for a broader range of purposes?** States are using these types of measures to:
- **Hold schools accountable** for performance;
  - **Provide diagnostic information** and reviews to find the causes of underperformance;
  - **Target supports and interventions** in schools for the schools and young people who need them most;
  - **Provide data to inform policy decisions** aimed at raising performance and addressing student and school needs;
  - **Provide transparency** for school performance through public reporting; and/or
  - **Set improvement goals.**
- Q. **How do you make sure that non-academic indicators don't cloak academic problems?** States can use a two-step accountability model where key outcome indicators are used to determine whether schools are having a problem and, in these cases, additional indicators are used to identify specific problems and plan school improvement. Note: This is also less expensive, since all measures are not collected for all schools.
- Q. **What are the major advantages of multiple measures?** Employing a broader range of measures can allow for more accurate assessment of school outcomes, promote more valid inferences about school performance, and provide more balanced incentives to teachers and principals. It also provides more context and information on which policymakers and state education leaders can base decisions about funding and strategies to improve struggling schools. Multiple measures can:
- **Help Maintain and Promote Strong Accountability.** Using more than one measure provides more accurate and consistent information about desired outcomes that helps to reduce the noise and distortions that affect any single measure and mitigates ways of gaming the system. Additional measures like chronic absenteeism and student mobility can be structured to reduce the incentive to push some students out of the system to raise test scores. Equally significant, identifying a combination of indicators for school success also helps schools and districts focus the system on achievement rather than test prep, and desired outcomes rather than a few narrow proxies for those outcomes.
  - **Promote State/Local Flexibility and Judgment.** States have unique priorities and needs. Multiple measures can be introduced in a way that will allow states to set their own agendas rather than being handcuffed by a narrow set of federal measures. This will ensure that accountability is not overly prescriptive and does not stifle innovation or limit the autonomy of state and local education agencies.
  - **Strengthen School Improvement Systems.** It is possible to have a multiple measures system that focuses on academic outcomes for accountability purposes while drawing on other broader measures in order to more effectively improve schools. By providing additional information and context for the unique situations each school faces, for example, states and districts can introduce improvement systems that better reflect the realities that teachers and school districts face every day. With a broader range of indicators, states and local agencies can make better decisions about how to address struggling schools and identify targeted interventions specific to the improvement needs of these schools. The broader set
-

---

of measures not only provides room for rigorous measures of college- and career-readiness, but also addresses equity concerns in current accountability measures. Depending on how they are structured, indicators of student growth, for example, provide a clearer picture of low-income schools' performance than only absolute measures of proficiency as is currently used under NCLB.

- Q. There must be hundreds of indicators that could be used. Which are the most common, and which are the main areas that would be most useful for policy decisions?** The RAND study organized the broad array of measures in five categories and, based on its review, encourages states to develop measures in these areas:

**Outcome Measures Focused on Academic Progress**

1. **Achievement and attainment beyond math and English Language Arts**, including measures of student growth, performance in other subject areas, and advanced course-taking.
2. **Progress indicators** demonstrating student advancement toward high school graduation or college- and career-readiness.

**Measures of Other School Factors and Outcomes Associated with School Success**

3. **Positive school culture**, including student and teacher satisfaction, academic challenge, engagement, safety, and orderliness.
4. **Positive behavioral, emotional, and physical health outcomes**, including indicators of attendance, suspensions, expulsion, and physical health.
5. **Information about discrepancies in resources** available to schools and students.

- Q. What does the research say about which measures work best or how best to implement them?** The research is unclear on a wide range of critical issues, including which and what combination of measures best address states' needs for accountability, transparency, and information about school performance. There is also a dearth of information about how much such a system might cost. But the evidence is clear about the limitations of using only a few specific subject areas to measure performance and suggests that there is a better way. There is growing consensus that the federal government should require states to have initial outcome determinations based at least on improved reading and math assessments (aligned to college- and career-ready standards) and accurate graduation rates. But federal law also should permit and encourage each state to propose and use additional, valid and reliable outcome measures, so long as their use is subject to rigorous peer review and research. Furthermore, federal law should require states to use diagnostic measures to review the lowest performing schools (perhaps the bottom five percent) and generally encourage and support states to experiment with a broader range of measures. While this might pose some challenge to states at a time when they have limited technical expertise and financial capacity, development of new measures could be accomplished by pooling efforts through state consortia or by prioritizing the most expansive set of measures for deeper diagnostic review for only the lowest performing schools (perhaps the bottom five percent).

- Q. This is not a time for heavy-handed government action or more unfunded mandates on cash-strapped states. How can the federal government best make this happen?** Perhaps the most powerful incentive is to allow states to use additional outcome measures for AYP, not as an additional all-or-nothing measure,

---

but as part of a weighted, blended measure. Reauthorization of ESEA could include incentives to develop state capacity and encourage coordination across states. Recognizing that states vary in their capacity to develop and test new measures, federal incentives could be leveraged to help states save money by encouraging them to coordinate their efforts through interstate consortia or other partnerships. Because we need to know more about what works most effectively, the federal government might consider sponsoring a competition similar to Race to the Top that spurs the development and evaluation of additional school performance measures. Finally, if federal law were to require diagnostic measures only for the lowest performing schools, the burden of federally required measures would be minimized.

- Q. Won't the use of multiple measures in reauthorized legislation, as in NCLB, enable states to game the system by setting a low bar for schools? What can be done to ensure that measures are adequately rigorous and comparable across states?** Experience with AYP shows that there is a need to permit and encourage states to broaden their measures if we want to achieve our education goals. AYP can serve as a floor, but as long as states are willing to meet core requirements (e.g., annual accountability determinations based in part on accurate assessments, graduation rates, etc.), they should be allowed to add additional measures that are designed to enhance validity and utility of accountability determinations. These measures would have to be approved through rigorous peer review, with additional evidence of validity evaluated over time.
- Q. How can we assure that these measures are valid and fair across districts?** States should also be required to assess the technical quality and effects on student outcomes when they add new measures to their accountability framework.
- Q. How do these recommendations align with what the Council for Chief State School Officers task force is recommending for next-generation accountability?** What the RAND study found supports CCSSO's approach. The task force has put a sensible proposal before state policymakers to encourage the federal government to empower states to develop a new model of school and district accountability that would use meaningful multiple measures and would be introduced over time. The new model would have states propose how they will measure progress toward college- and career-readiness and determine school success based on performance of all students through a commonly shared measure of graduation rates and improved and more rigorous state assessments that include growth measures. The CCSSO task force also supports multiple measures that can be used for broader purposes, including improving state- and school-level decision-making, providing more timely and transparent information, and ensuring that the lowest performing schools receive significant, meaningful interventions.
- Q. How can you use multiple measures in a way that is simple and doesn't drown users in a barrage of unfocused information?** One way is that states can combine measures into an index that assigns weights to different measures and provide guidelines on how different indicators are meant to be used within the overall school improvement and accountability process.
- Q. What are the risks and tradeoffs of implementing multiple measures?** Using these measures to drive accountability systems has complex implications. Lawmakers will need to carefully consider what measures are the most valuable and how they will be used in accountability decisions. Policy must strike an appropriate
-

---

balance in an array of areas, including:

- **Breadth and focus.** Lawmakers need to set priorities about what is important to measure and determine the benefits of assessing many crucial outcomes, as opposed to highlighting only a few areas.
- **Complexity and transparency.** More sophisticated measures might be difficult for educators and the public to understand, while simpler measures are easier to interpret but of less value for decision-making.
- **Comprehensiveness and affordability.** Although there is little evidence about what states currently pay to gather information, measuring many facets of school performance will undoubtedly cost more.
- **Uniformity and flexibility.** Establishing a uniform set of measures would allow for easy comparisons across states but could stifle innovation and local decision-making.

**Q. Do you have any suggestions about how best to address these trade-offs?**

The best suggestion we can offer is to learn from the past. The narrow focus of the current AYP measures needs to be broadened. The federal government should encourage state efforts to be transparent and useful for decision-making. It should recognize that states need some leeway in determining what they will measure, knowing that already more than half have moved in this direction.