

CED IN BRIEF

Policy Forum Highlights

Educational Assessment and Accountability

CED's policy statement *Measuring What Matters*¹ emphasizes the central importance of assessment and accountability in efforts to reform American education and improve student learning. Its key recommendation: tests should be used and improved now, rather than resisted until they are perfect, because they provide the best means of charting our progress toward the goal of improved academic achievement.

President George W. Bush has proposed changes in federal education law that echo these themes. On April 3, 2001, CED's report served as the backdrop for a luncheon speech by **Senator James Jeffords** (R-VT) and a panel discussion on the federal role in assessment and accountability with **Gordon Ambach**, Executive Director of the Council of Chief State School Officers; **Christopher Cross**, President of the Council for Basic Education; and **Michael Feuer**, Executive Director of the National Research Council's Center for Education. This issue of *CED in Brief* highlights key points from the presentations.

The program began with remarks by **Senator Jeffords**, chair of the Senate Committee on Health, Education, Labor, and Pensions. Senator Jeffords is a sponsor of the Better Education for Students and Teachers (BEST) bill. BEST embraces key elements of President Bush's proposals, calling for annual testing of all students in third through eighth grades in reading and math, using state tests that are aligned with state academic standards. Results from the National Assessment of Educational Progress (NAEP), which is currently administered on a voluntary basis in grades 4, 8, and 12, would be used to "confirm" state results for the purposes of providing federal rewards for states making academic progress and determining which states should lose some

federal funds as a sanction for poor performance. (Thus administration of NAEP would be mandatory in all states.) Schools that consistently underperform would be required to give students the option to attend other public schools and allow parents to use federal funds for supplemental academic assistance for their children. New federal funds would be made available to help the states develop new assessments and to reward states that meet performance objectives based on criteria to be set by the Secretary of Education.

SENATOR JEFFORDS:

Congress is in the midst of rewriting the major federal law affecting elementary and secondary education, which will guide federal education policy for most of this decade. The BEST proposal makes perhaps the most dramatic changes in this policy since federal education programs were first created in 1965.

You should like BEST because it measures up to CED's excellent and timely report, *Measuring What Matters*. BEST combines an increased emphasis on accountability with additional resources for helping all students, schools, districts, and states meet high standards. New assessments must be in place by the 2005-2006 school year. These tests must be based on rigorous content standards and have challenging performance standards. All students, including special needs students, must take the same tests. BEST also requires that state assessments meet recognized professional standards for reliability, validity, and fairness. We are building on our current assessment technology to increase the accountability of states, districts, and schools for helping our children to achieve. Rewards will accompany improved performance, and sanctions will follow failure to achieve.

What we've done in writing this legislative language is only the start of a very difficult process of change and improvement. In this regard, I would like to focus on one of CED's recommendations: "A performance-based educational system built on measuring student achievement requires new investment. It can't be constructed on the cheap." I agree with this completely. What the President has asked for and what we have done in BEST can lead to unintended negative consequences if we do not implement new law wisely and with care. To do that will take new investments. We need more research and development on assessment technologies. We need to give states and localities the resources and time to implement a high quality assessment system.

Finally, we need to ensure that there are adequate resources available to make the changes in schools that will allow every child to move ahead. As CED's report points out, measuring performance is only a means to the larger end of school improvement. Assessment and accountability won't by themselves make education better. We must create and support a system of care for pre-school age children that ensures that they all arrive at kindergarten ready to learn. Once in school, every child needs to experience well-prepared teachers, high standards for learning, state of the art assessments, systems that support continuous improvement, and access to continuing education and training after formal schooling ends.

Following Senator Jeffords' speech, CED Vice President and Director of Education Studies **Janet Hansen** moderated a panel discussion with **Gordon Ambach**, **Christopher Cross**, and **Michael Feuer**. Hansen pointed out that the Bush and BEST proposals represent an unprecedented level of federal involvement in an enterprise which has, up until now, been focused largely on state-level reforms. She asked the panelists to respond to the challenge posed by these proposals: How can Washington be a constructive influence on continuing state and local efforts to use assessment and accountability to improve student learning? *(The following opinions are those of the speakers and do not necessarily represent the views of CED Trustees.)*

GORDON AMBACH:

My organization has been a leader in advancing state assessments and has strongly supported the idea of having incentives in programs. Nevertheless, I have reservations about BEST's call for annual testing, about the proposed uses of NAEP, and about whether the proposed system of rewards and sanctions can in fact work.

Currently, only about 13 states test every student annually in reading and math in grades 3 through 8. Under BEST, 37 states will have to substantially change their assessment systems. The burden on the states could be alleviated by allowing states to meet BEST's requirements by using a mixture of state and local tests.

NAEP should be kept on a voluntary basis and used every other year for the purpose of determining trend lines. Using NAEP for high-stakes financial decisions and trying to press it into annual use for this purpose raises major questions, since the tests were not designed with this use in mind. A critical issue is: What does it mean to say that NAEP scores will be used to "confirm" the results on state tests?

With respect to rewards and sanctions, incentives are really the important thing. The current proposals leave most of the decisions about how the rewards would work to the Secretary of Education's discretion. The sanctions would penalize states by taking administrative money away from the state level. We'd prefer to emphasize incentives: for example, adding a percentage to the Title 1 appropriation that could be allocated among states based on whether they met performance targets that they themselves set. That would be an incentive for everyone to work on meeting the state's goals, and it avoids problems because it allows each state to focus on its own targets using its own assessments.

It's important to remember that state testing systems haven't been designed for high-stakes financial decisions even within the states, and NAEP was never intended to be used for this purpose. It's risky to start using these instruments in these new ways without a lot of analysis of whether and how this can work and whether the technology is there to make rational, fair decisions. We have urged Congress and the Administration to fund a National Academy of Sciences study for two years to evaluate the feasibility of using state tests and NAEP for high-stakes financial decisions and, if feasible,

to design procedures for the Secretary to determine awards.

CHRISTOPHER CROSS:

CED's report makes an important contribution by acknowledging the complexities of assessment and accountability. I applaud its stress on the importance of paying attention to the complexities and the importance of making additional investments in tests and accountability systems. We need new models that are capable of meeting the demands now being made on them. It's crucial that these messages get through to policymakers not only in Washington but in the states as well.

One of the things I learned in seven years chairing the Independent Review Panel on Title 1 for the U.S. Department of Education is that it is painstaking work to do the kinds of things being contemplated here. The fact that many states have not yet met the assessment and accountability requirements enacted in the 1994 reauthorization of federal education law attests that this is tough work and that accountability provisions are not easy to deal with.

The "test overload" issue is an important one that should be examined with some care. Are schools using too many tests and overemphasizing testing to the detriment of education? We tend to create geologic layers of tests: We add on new ones without examining and perhaps eliminating the ones that are already in place. This contributes to public confusion: What do the results of all these differing assessments mean?

CED's report mentions the importance of alignment between what's taught and what's tested. This point has to be hit again and again. Some states still don't get it. Policy makers don't understand why they shouldn't use a generic, national norm-referenced test to measure students' performance on specific state academic standards.

The report also mentions that Americans strongly support testing, but they express concern about testing's limitations. We need to find ways of supplementing test data with other information to reassure the public and educators that we are using good measures of the breadth and depth of student performance.

The tension between timeliness and complexity needs additional attention. Complex assessments tend to take longer to report

results. We not only have to improve assessments, but we have to improve the infrastructure through which they are administered, scored, and reported. Capacity is a major issue. We're not producing testing experts fast enough to meet current demand, nor have we invested enough in increasing people's capacity to interpret and use assessment data.

Finally, there's a question of reliability: What assessments are reliable for what purposes and where can one get information about that?

MICHAEL FEUER:

I am actually rather positive about our growing ability to meet the significant challenges that have been identified here. Concerns about whether assessment and accountability systems are up to the demands being placed on them are not misplaced, but in recent years we have learned a great deal about some of the questions that are posed by the new federal initiatives.

The NRC has a new report called *Knowing What Students Know*² which is based on three years of research about how assessments can be better aligned with our best knowledge about cognition and how students learn and what makes for effective teaching. We are discovering ways of getting assessment on a better course to contribute to educational improvement.

Nevertheless, there are still dilemmas that the current federal initiatives or any assessment program must confront. Parents and the public are asking a simple question: "How are our kids doing?" It's hard to provide a simple answer, though, because both the apparatus of assessment and what it means to be educated are complex. There's consensus in the professional community that multiple measures are important. It is inappropriate to rely on a single score to make decisions about anything of consequence. But with multiple measures comes the possibility of mixed signals about performance and concerns about testing overload.

Multiple measures often produce discrepant results. A popular idea is to put all the results on one scale. Experts have advised the NRC, however, that this is not really possible. So the potential for public confusion remains.

We also have to resolve the dilemma of whether tests are intended as tools of measurement or tools of change. Education is one of the few areas where we are trying to use the measurement apparatus to effect change. Tests that are used as instruments of change, however, may lose their value as instruments of measurement. This may strengthen the argument for multiple measures and may indicate that we will have to settle for occasional discrepancies. I actually think that we should give the American public credit for being able to figure out that tests are given under certain conditions with certain objectives in mind and with certain kinds of content emphases and that they therefore produce certain kinds of results. Other tests given for other purposes with different emphases may well produce different findings.

How much should we worry about assessments and accountability being used in inappropriate ways? There's room for optimism here. We know a lot about what validity, reliability, and fairness mean when these terms are applied to tests or test uses. Senator Jeffords' comments reflect the fact that policy makers are coming to understand the importance of these criteria. It's true, though, that accountability systems can create incentives to use tests and test results in inappropriate ways. We aren't going to know all we need to about how to get this right until we do more research about incentives and their effects.

It's also important to acknowledge that to date the best tests we know of only provide an estimate of one portion of learning. We should all keep in mind that we're using an estimation technology here. Moreover, we're trying to make complicated judgments and decisions based on a technology that doesn't have the fine-tuning knob attached to it yet. But, if Senator Jeffords gets this right, and we get the time and the resources to really think through these things, I am optimistic that we are on the cusp of a big revolution in the appropriate use of assessment to make significant improvements in the way children learn and the way they are taught.

CED President **Charles Kolb** concluded the forum by harking back to the first summit on education between the President and the

nation's governors at Charlottesville, Virginia in 1989.

CHARLES KOLB:

I would like to close this forum by siding with the optimists. The kind of debate we are having here and in Congress right now and around the country about how to improve education is exactly what was envisioned in Charlottesville. That's the basis for my optimism. We don't know where the debate will come out, but this type of discussion is a significant step in the right direction. Many thanks to our panelists and guests.

¹ CED's report, *Measuring What Matters: Using Assessment and Accountability to Improve Student Learning*, can be viewed and downloaded at CED's website: www.ced.org Ordering information is also available at that site or by calling 212-688-2063.

² The National Research Council's report, *Knowing What Students Know*, is available from the National Academy Press. Ordering information can be found at www.nap.edu or by calling 888-624-8373 or 202-334-3313.

CED is a nonprofit, nonpartisan organization of over 220 business leaders and university presidents. Since 1942, its research and policy programs have addressed the nation's most pressing economic and social issues, including education reform, workforce competitiveness, campaign finance, health care, and global trade and finance. CED promotes policies to produce increased productivity and living standards, greater and more equal opportunity for every citizen, and in improved quality of life for all.

Committee for Economic Development

**477 Madison Avenue
New York, NY 10022
Phone: 212-688-2063
Fax: 212-758-9068**

**2000 L Street NW, Suite 700
Washington, DC 20036
Phone: 202-296-5860
Fax: 202-223-0776**

Web Site: www.ced.org