

Richard G. Innes
2836 Deerfield Drive
Villa Hills, KY 41017

November 10, 2009

RE: NAEP Testing for Students with Disabilities and English Language Learners

The impact of exclusion and inclusion of students with learning disabilities on the validity of National Assessment of Educational Progress (NAEP) results is a very important subject, but it is far from a new problem. This threat to NAEP validity has been a concern for at least decade.

Open, public concerns about exclusion of learning disabled students were first raised in March 1999 following my discovery that Kentucky's comparatively large increase in fourth grade NAEP reading scores between 1994 and 1998 were accompanied by an equally large increase in the percentage of the raw sample that was excluded in 1998 due to learning disabilities.

Simple regression analysis indicated that at least half – and possibly all – of Kentucky's score rise could be explained by the fact that exclusion of learning disabled students had skyrocketed from four percent in 1994 to ten percent of the raw sample NAEP pulled for testing in the state in 1998. The threat to the validity of the NAEP was very obvious. Could we be confident in NAEP's message that Kentucky was one of the most improved states on fourth grade reading in 1998, or was the state merely the beneficiary of a lack of control over exclusions?

Ten years later, while NAGB still endeavors to find answers to that question, I am pleased to see this needed effort has indeed continued despite the complex challenges involved. Certainly, the new report from the Technical Advisory Panel (TAP) on Uniform National Rules for NAEP Testing of Students with Disabilities looks promising, and I welcome a chance to offer some thoughts on this evolving process.

NOTE: Richard Innes is education policy analyst for the Bluegrass Institute for Public Policy Solutions, of Bowling Green and Lexington, KY. The comments are his own and have not been reviewed by the institute. INTERNET70224.434@compuserve.com.
Phone: 859-344-0406; Cellular 859-466-8198

RE: TAP Recommendation 1.

Encourage as many students as possible to participate in NAEP, and provide for the use of allowable accommodations that are necessary to enable students with disabilities to participate.

I support these TAP recommendations with one important caution – “allowable accommodations” must include more than the concept of enabling students to participate. This must also be driven by the equally important concept that the accommodations must not alter the construct *nor artificially inflate scores*.

That drives a suggestion that I don’t see in the TAP recommendations. There needs to be a plan for scientific split sample testing of any proposed or currently used testing accommodations to determine if they are truly construct neutral.

As a bonus, properly designed testing could provide high confidence correction factors to weigh results when accommodations are deemed desirable even though they do incur some degree of consistent scoring inflation.

This scientific study will be involved and probably involve some notable costs, a factor which will impact my following comments on some other TAP proposals which seem to offer lower “bang for the buck” potential.

RE: TAP Recommendation 2.

Clarify and expand NAEP’s guidance to schools, encouraging maximum participation of students with disabilities.

These are very good suggestions.

RE: TAP Recommendation 3.

Report separately on NAEP results for IEP and 504 students.

These are also good recommendations. The TAP’s sensitivity to guarding the trend lines is particularly noteworthy.

RE: TAP Recommendation 4.

Provide incentives for schools to include students with disabilities.

Of the two suggestions to incentivize schools to maximize participation, the second one seems more reasonable. I would go with the plan to develop “a uniform SD participation guideline based on a percentage of the total student population, regardless of the percent identified as SD.”

There is great variation of identification of students with disabilities, and these variations appear, as the TAP points out, to be largely due to policy decisions rather than true variations in the incidence of real disabilities in different states. One state should not be placed at a disadvantage to another on the NAEP due to these policy decisions. In fact, there is concern that some of these policy decisions appear to actually be counter to good education and are more directed at securing larger amounts of support money for students who actually are not learning disabled, but are suffering from poor instruction. NAEP should not inadvertently adopt policies that would aid and encourage such inappropriate behavior.

RE: TAP Recommendation 5.

Support research efforts to develop targeted testing for all students at both the top and bottom levels of achievement, with sound procedures to identify students to receive targeted test booklets on the basis of their performance on some standard indicator of achievement.

I view this as the most potentially problematic of the TAP recommendations. It is much like the concept of branching computer-based testing, but it will be done, apparently, without the aid of technology. That raises all sorts of problems.

There are no easy ways to select which students get which level of “branched” NAEP tests.

I think it would be highly problematic to try to use state test results to determine the NAEP branch. Recent NCES research shows state tests now vary dramatically in difficulty. A lot of high-scoring students in one state might not come close to matching top students in another state that has much more rigorous standards.

Furthermore, at least in Kentucky, my research shows that the state test scoring has constantly been getting easier over time.

Given these variables, how would NAGB ever confidently decide what cut scores to use with each of 50 different state assessments to determine the appropriate NAEP branch level test to give each student?

Would NAEP validity suffer if too many students in a state got too difficult a test, or too easy a test, as a consequence of this automatic, 50-different-outcomes threat to NAEP validity?

Basically, using the state tests to branch NAEP would be highly problematic, and probably almost impossible to defend.

Given that using state tests is obviously a very bad choice, implementing this “branching NAEP” idea would require two rounds of NAEP conducted tests, probably administered on different days with a delay for scoring in between.

That is a logistic nightmare.

Extended testing will generate huge pushback from parents who are already upset about the amount of testing their kid's experience. Schools won't be pleased, either.

There are a number of other problems with this proposal, but I'll just touch on a few more.

Will the possible benefits of "branching NAEP" results outweigh the considerable costs?

Also, the design of these "branching-like" tests would be complex, adding more costs.

Very simply, this is not a very attractive idea. I'd rather see the money spent on decent split sample testing to really establish the validity of accommodations and any score corrections that their use should require.

RE: TAP Recommendation 6.

Encourage and review research on the identification and progress of students who have a significant cognitive disability but in the short term do not test this 1% of students on NAEP.

I agree with this cost-cognizant approach. The number of students involved is very small and getting statistically valid scores for this small group would be a challenge in any event.

RE: TAP Recommendation 7.

Assess the English language proficiency of students with disabilities drawn for the NAEP sample and provide NAEP-approved, linguistically appropriate accommodations for them before determining whether additional accommodations may be needed to address any disabilities these students may have.

This is another good, common-sense idea. Let's determine if a student just has a challenge working in English before we assume his or her low performance is due to a true disability. Certainly, we need to include these students as they have become a significant portion of the school population in a number of states.

That leads to another issue that NAGB needs to discuss which isn't in the TAP report.

Added item to consider – How do changing demographics impact NAEP interpretation?

Making valid state to state comparisons with NAEP results is now being seriously compromised by the huge changes in demographics that have occurred in some states since state NAEP began in 1990. I believe the demographic impacts far outweigh the impacts of variation in exclusion rates in some cases.

For example, this table shows how California's fourth grade reading scores might have looked if the state's student demographics had remained stable at the 1992 levels.

**Comparison of California NAEP Grade 4 Reading Scores
Scores Assuming Constant 1992 Student Demographics
Vs. Scores Officially Reported**

Year	Weighted Average Score for All Students with Demographics Held Constant to 1992	Reported California NAEP Scores
1992 ¹	202.3	202
1994 ¹	197.3	197
1998	203.5	202
2002	211.4	206
2003	211.9	206
2005	213.4	207
2007	216.2	209

¹ Accommodations were not permitted for this assessment

Note that the official score change in California was only 7 points, but if the demographics had stayed constant, the score rise would be twice that amount, a very serious difference given the relatively flat reading performance of all the states.

California has been rather severely criticized over the years for its low performance on the NAEP, but the data in the table above indicates that the state may actually have made much more notable progress than NAEP is showing. That is a validity issue that needs NAGB attention.

I have provided an Excel spreadsheet used to calculate the constant demographics scores for California for those who are interested.

In closing, I am pleased that NAGB is working on important issues of NAEP validity. We need a solid NAEP to help the country determine the true rate of progress in education in the country and to inform the states as they attempt to improve their own assessment programs. Any issues that detract from the validity of NAEP results are therefore of major importance to everyone throughout the country, including those of us in Kentucky where some of the issues first came to public attention.

Sincerely yours,

Richard G. Innes